REMARKS

Claims 39-59 were under examination as of the issuance of the Office Action. Claims 47-53, 57 and 59 are withdrawn in response to a Restriction Requirement set forth in the Office Action, but are capable of rejoinder in accordance with the provisions of MPEP § 821.04. Claims 39-46 stand rejected. In the Amendment to the Claims spanning pages 2-5 of this paper, claims 39, 43, 44, 57 and 59 have been amended, new claim 60 has been added and claims 54-56 and 58 have been cancelled without prejudice. Accordingly, upon entry of the amendments presented herein, claims 39-53, 57, 59 and 60 will remain pending in this application.

Support for the amendments to the claims and for the new claims may be found throughout the specification and in the claims as originally filed. Support for the amendments to claim 39 can be found throughout the specification, for example, at page 25, lines 10-24, at page 25, lines 25-38 and at page 3 of Table 1, and in the claims as originally filed. Support for the amendments to claims 43, 44 and 59 can be found throughout the specification, for example, at page 34, line 28 to page 36, line 30, and in the claims as originally filed, for example, claim 44. Claim 57 has been amended merely to attend to formalities, *i.e.*, the correction of claim dependency.

New claim 60 has been added to cover a particular embodiment of the present invention. Support for new claim 60 can be found throughout the specification, for example, at page 25, lines 10-24.

No new matter has been added by the claim amendments or the introduction of the new claim. The amendments to and/or cancellation of the claims should not be construed as an acquiescence to the validity of the Examiner's rejections and were done solely in the interest of expediting prosecution and allowance of the claims. Applicants reserve the right to pursue the claims as originally filed in one or more further applications.

Restriction Requirement and Election

Claims 39-59 are subject to a restriction requirement. The Examiner has required restriction, under 35 U.S.C. § 121, between the following inventions in the above-identified application:

Group I: Claims 39-46, drawn to an isolated nucleic acid molecule, vector, host cell, and method for making a polypeptide, classified in class 435, subclass 69.1;

Group II: Claims 47-53, drawn to a method for making a fine chemical, classified in class 435, subclass 41;

Group III: Claims 54-56, drawn to an isolated polypeptide, classified in class 530, subclass 350;

Group IV: Claim 57, drawn to a method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject comprising detecting the presence of at least one nucleic acid molecule, classified in class 435, subclass 6;

Group V: Claim 58, drawn to a method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject comprising detecting the presence of at least one polypeptide, classified in class 435, subclass 7.1;

Group VI: Claim 59, drawn to a host cell comprising a nucleic acid molecule that is disrupted, classified in class 435, subclass 252.3.

Applicants hereby confirm the provisional election of Group I (claims 39-46, drawn to an isolated nucleic acid molecule, vector, host cell, and method for making a polypeptide) as initially elected by the undersigned during a telephone conversation with the Examiner on January 4, 2006, ***without traverse*** for continued examination in the present application. Applicants' election of the foregoing subject matter is without prejudice to Applicants' rights to pursue non-elected subject matter in other applications. Furthermore, Applicants reserve the right to traverse the restriction between the non-elected groups in this or a separate application.

With regard to the remaining withdrawn claims, it is Applicants' understanding that once a composition claim is found to be allowable, the pending claims that depend from or otherwise include all the limitations of an allowable composition claim will be rejoined in accordance with the provisions of MPEP § 821.04. Accordingly, Applicants respectfully request rejoinder of claims 47-53, 57 or 59 should any of the pending composition claims be found to be allowable.

Priority

Applicants note that certified copies of the foreign German patent applications will be filed shortly and certainly prior to the issuance of a patent based on the above-identified patent application.

Rejection of Claims 39-46 Under 35 U.S.C. § 112, Second Paragraph

The Examiner has rejected claims 39-46 under 35 U.S.C. § 112, second paragraph as "being indefinite for failing to particularly point out and distinctly claim the subject matter which applicants regard as the invention."

Specifically, the Examiner is of the opinion that "the phrase 'set forth' renders the claim vague and indefinite because it is unclear if applicants are actually referring to the specific SEQ ID NOs" (Office Action, page 5). In the interest of expediting examination, but in no way acquiescing to the validity of the Examiner's rejections, Applicants have amended claim 39 as recommended by the Examiner, *i.e.*, to read as "the nucleotide sequence of SEQ ID NO:179" or "the amino acid sequence of SEQ ID NO:180."

The Examiner is also of the opinion that "the phrase 'a complement thereof' renders the claims vague and indefinite because it is unclear if applicants are referring to a nucleotide sequence that is the full and complete complement of SEQ ID NO: 179 or a complement of a part of SEQ ID NO: 179" (Office Action, page 5). In the interest of expediting examination, but in no way acquiescing to the validity of the Examiner's rejections, Applicants have amended claim 39 as recommended by the Examiner, *i.e.*, to read as "the complement thereof."

Applicants submit that the foregoing claim amendments render this rejection of claim 39, and the claims depending therefrom, moot. Accordingly, Applicants respectfully request reconsideration and withdrawal of the rejection of the pending claims under 35 U.S.C. §112, second paragraph.

Rejection of Claims 39-46 Under 35 U.S.C. § 112, First Paragraph (Written Description)

The Examiner has rejected claims 39-46 under 35 U.S.C. § 112, first paragraph as "failing to comply with the written description requirement." In particular, the Examiner is of the opinion that

> [c]laim 39 is drawn to a genus of complements of SEQ ID NO:179, a genus of complements of any nucleic acid molecule encoding SEQ ID NO:180, a genus of nucleic acid molecules encoding naturally occurring allelic variants of a polypeptide comprising SEQ ID NO:180, a genus of nucleic acid molecules that is at least 50% identical to SEQ ID NO:179, and a genus of nucleic acid molecules comprising any fragment of at least 15 contiguous nucleotides of SEQ

> ID NO:179. The scope of... each genus includes many members with widely differing structural, chemical and physiochemical properties including widely differing nucleotide sequences. Furthermore, each genus is highly variable because a significant number of structural differences between genus members exist. (Office Action, pages 7-8)

Applicants respectfully traverse this rejection on the grounds that, based on the teachings in Applicants' specification and the general knowledge in the art at the time of the invention, one skilled in the art would reasonably conclude that the Applicants were in possession of the claimed invention at the time the application was filed. In the interest of clarity, Applicants will address each aspect of this rejection below.

Rejection of Claims Directed to Complements of Nucleic Acid Molecules

As set forth above, the Examiner asserts that the claimed genus of complements of SEQ ID NO:179 and the claimed genus of complements of any nucleic acid molecule encoding SEQ ID NO:180 "is not described in the specification in such a way to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention" (Office Action, page 7).

Applicants respectfully traverse this rejection on the grounds that one skilled in the art would reasonably conclude that the Applicants were in possession of the claimed invention at the time the application was filed. However, in the interest of expediting examination and in no way acquiescing to the validity of the Examiner's rejection, claim 39 has been amended to be directed to ***the complement*** of the recited nucleic acid molecule. Applicants submit that this amendment renders the Examiner's rejection moot.

Rejection of Claims Directed to Nucleic Acid Molecules Encoding Allelic Variants

The Examiner also alleges that the genus of naturally occurring variants of claim 39 is not adequately described in view of the disclosure of "only one member of this genus." In particular, the Examiner is of the opinion that

> [t]he nature of allelic variants is that they are variant structures where the structure and function of one does not provide guidance to the structure and function of others. Although the specification discloses only one allele within the scope of the genus which is SEQ ID NO:179, the general knowledge in the art concerning alleles... is that there is no any [sic] indication of how the structure of one allele is representative of other unknown alleles having concordant or discordant functions. (Office Action, page 8)

Applicants traverse this rejection on the grounds that, based on the teachings of the specification and the state of the art at the time of filing of the present application, one skilled in the art would reasonably conclude that Applicants were in possession of the genus of "naturally occurring variants" of polypeptides comprising SEQ ID NO:180 at the time the application was filed. Applicants respectfully disagree with the Examiner's characterization of the term "allelic variant" as described in the specification and known in the art. Applicants direct the Examiner's attention to page 28, lines 12-18 where Applicants define allelic variant as follows:

> In addition to the *C. glutamicum* SMP nucleotide sequences shown in Appendix A, it will be appreciated by those of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of SMP proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the SMP gene may exist among individuals within a population due to natural variation... Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the SMP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in SMP that are the result of natural variation and do not alter the functional activity of SMP proteins are intended to be within the scope of the invention.

Applicants further note that patents, issued at the time of the filing of the present application, describe such naturally occurring variants in a consistent manner. Applicants direct the Examiner's attention to, for example, U.S. Patent No. 5,882,893, issued on March 16, 1999, in which independent claims 1, 14 and 15 are directed to "allelic variants." The specification of U.S. Patent No. 5,882,893 characterizes this term as follows:

> In addition to the mACHR-6 nucleotide sequence shown in SEQ ID NO:1 or 4, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of mACHR-6 may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the mACHR-6 gene may exist among individuals within a population due to natural allelic variation... Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the mACHR-6 gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in mACHR-6 that are the result of natural allelic variation are intended to be within the scope of the invention. Such allelic variation includes both active allelic variants as well as non-active or reduced activity allelic variants, the later two types typically giving rise to a pathological disorder. (Column 18, lines 3-32).

In view of the foregoing teachings in Applicants' specification and the art, Applicants submit that a skilled artisan would appreciate that the recitation of a specifically defined nucleotide sequence encoding a polypeptide is representative of the larger genus of nucleic acid molecules encoding naturally occurring allelic variants with similar structure and function.

Accordingly, one skilled in the art would conclude that the Applicants were in possession of the claimed naturally occurring allelic variants at the time of filing of the present application.

Notwithstanding the foregoing, in order to expedite examination, but in no way acquiescing to the validity of the Examiner's rejection, Applicants have cancelled claim 39(c) directed to nucleic acid molecules encoding naturally occurring allelic variants of a polypeptide comprising the amino acid sequence of SEQ ID NO:180, thereby rendering the rejection of this claim, and the claims depending therefrom, moot.

Rejection of Claims Directed to Sequences of 50% Identity to the Nucleotide Sequence of SEQ ID NO:179

The Examiner has further asserted that the genus of nucleic acid molecules that is at least 50% identical to SEQ ID NO:179 "is not described in the specification in such a way to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention" (Office Action, page 7).

Applicants respectfully traverse the foregoing rejection for the following reasons. Applicants would like to direct the Examiner's attention to Example 14 of the *Revised Interim Written Description Guidelines Training Materials*. This example provides that a claim directed to variants of a protein having SEQ ID NO:3 "that are at least 95% identical to SEQ ID NO:3 and catalyze the reaction of A→B" with an accompanying specification that discloses a single species falling within the claimed genus, satisfies the requirements of 35 U.S.C. §112, first paragraph for written description. The rationale behind the foregoing conclusion, as presented by the *Written Description Guidelines*, is that "[t]he single species disclosed is representative of the genus because all members have at least 95% structural identity with the reference compound and because of the presence of an assay which Applicant provided for identifying all of the at least 95% identical variants of SEQ ID NO:3 which are capable of the specified catalytic activity." The Guidelines also provide that "***[t]he procedures for making variants of SEQ ID NO:3 are conventional in the art and an assay is described which will identify other proteins having the claimed catalytic activity. Moreover, procedures for making variants of SEQ ID NO:3 which have 95% identity to SEQ ID NO:3 and retain its activity are conventional in the art.***"

Similarly, in the present case, claims 36(d) and 60, and claims depending therefrom, as amended, are directed to nucleic acid molecules comprising a nucleotide sequence which is at

least 90% or 95% identical to SEQ ID NO:179 and which code for polypeptides that have a phosphoenolpyruvate carboxykinase activity. The indication in Example 14 of the Written Description Guidelines that the production of polypeptides which contain a 5% variation from a specific sequence is routine in the art can be equated with the production of nucleic acid molecules which contain a 5% variation from a specific sequence. Furthermore, Applicants have disclosed in the instant specification assays for identifying all of the nucleic acid molecules of at least 90% or 95% identity to SEQ ID NO:179 having a phosphoenolpyruvate carboxykinase activity (see, for example, Examples 4-9 at page 51, line 29 to page 57, line 24 of the specification and Example 11 at page 58, line 24 to page 60, line 6 of the specification).

Accordingly, for at least the foregoing reasons, it would have been clear to one skilled in the art that Applicants were in possession of the claimed invention at the time the application was filed. Applicants, therefore, respectfully request reconsideration and withdrawal of the rejection of claim 36(d), and claims depending therefrom, under 35 U.S.C. § 112, first paragraph as lacking written description.

Rejection of Claims Directed to Nucleotide Sequences of 15 Contiguous Nucleotides

The Examiner asserts that the genus of nucleotide sequences of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179 is not adequately described such that one skilled in the art would reasonably conclude that the Applicants were in possession of the genus. In particular, the Examiner is of the opinion that

> [w]hile the specification discloses a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NO:179 which encodes a phosphoenolpyruvate carboxykinase from *Corynebacterium glutamicum* consisting of the amino acid sequence of SEQ ID NO:180; there is no recitation of any particular structure to function relationship in the claims which would define any biological properties and enzyme activities common to the members of each genus. Furthermore, the specification does not define any structure to function relationship for each claimed genus other than the polynucleotide of SEQ ID NO:179 encoding a phosphoenolpyruvate carboxykinase consisting of the amino acid sequence of SEQ ID NO:180. (Office Action, page 8)

Applicants respectfully traverse the rejection on the following grounds. Applicants submit that based on the identification of the nucleotide sequence of SEQ ID NO:179 as set forth in the present application, one skilled in the art would be able to readily envision all fragments of at least 25 contiguous nucleotides in length that could serve as, for example, probes, primers, antisense molecules or enzymatically active fragments. Indeed, the specification teaches that

nucleotide sequences of at least 25 contiguous nucleotides of SEQ ID NO:179 can be used as probes and primers (see page 25, line 25 to page 26, line 9 of the specification) in addition to use as antisense molecules (see page 31, line 16 to page 33, line 9 of the specification). Such fragments do not necessarily need to exhibit a phosphoenolpyruvate carboxykinase activity. Accordingly, the recitation of "a particular structure to function relationship... [so as to]... define biological properties and enzyme properties common to members of the genus," as the Examiner asserts is required, is, in fact, unnecessary. Instead, the ability of the claimed nucleotide sequences to hybridize to strands of RNA or DNA, without necessarily exhibiting any enzymatic activity, renders the claimed nucleic acid molecules operative for purposes of the invention.

Moreover, Applicants submit that the present specification provides teachings sufficient to demonstrate that Applicants were in possession, at the time of filing, of the genus of fragments of at least 25 contiguous nucleotides of SEQ ID NO:179 which are, in fact, capable of exhibiting phosphoenolpyruvate carboxykinase activity. At page 26, line 33 to page 27, line 5 of the specification, Applicants teach that such fragments could be designed by the incorporation of active domains of the phosphoenolpyruvate carboxykinase protein that participate in the metabolism of carbon compounds such as sugars, or in energy-generating pathways. Such domains were well known in the art at the time of the filing of the present application. For example, the domain [F/Y]-P-S-[A/G/M/S]-C-G-K-T-[N/S] (amino acid residues 270-278 of SEQ ID NO:2 encoded by nucleotide residues 908-934 of SEQ ID NO:1) had been identified at the time of filing of the present application as a highly conserved domain among phosphoenolpyruvate carboxykinase proteins amongst various organisms including prokaryotes, eukaryotes and archaebacteria (see PROSITE entry for PEPCK_GTP, attached herein as Appendix L; see also Lewis *et al.* "Cysteine 288: An Essential Hyperreactive Thiol of Cytosolic Phosphoenolpyruvate Carboxykinase (GTP)" (1989) *J. Biol. Chem.* 264(1):27-33), attached herein as Appendix M). As described by Lewis *et al.*, the cysteine residue, in particular, at the center of the aforementioned domain was known to be essential to the catalytic properties of phosphoenolpyruvate carboxykinase. Moreover, Example 8, at page 56, lines 1-28 of the specification, in addition to standard techniques known in the art at the time of the filing of the present application, provide means of identifying those fragments of at least 25 contiguous nucleotides which retain a phosphoenolpyruvate carboxykinase activity.

For each of the foregoing reasons, Applicants submit that sequences of at least 25 contiguous nucleotides of SEQ ID NO:179, including fragments which exhibit enzymatic activity and, further, those that do not exhibit such activity, are sufficiently described to demonstrate to a skilled artisan that Applicants were in possession of the claimed sequences at the time of filing. Applicants, therefore, respectfully request reconsideration and withdrawal of the rejection of the pending claims under 35 U.S.C. §112, first paragraph, as lacking written description.

Rejection of Claims 39-46 Under 35 U.S.C. § 112, First Paragraph (Enablement)

The Examiner has also rejected claims 39-46 under 35 U.S.C. § 112, first paragraph as lacking enablement. In particular, the Examiner is of the opinion that

> [t]he nature and breadth of the claim [39] encompasses any nucleic acid molecule comprising a nucleotide sequence which is at least 50% identical to SEQ ID NO:179 or a complement thereof, any nucleic acid molecule which encodes any naturally occurring allelic variant of a polypeptide comprising SEQ ID NO:180, and any isolated nucleic [acid] molecule comprising a fragment of at least 15 contiguous nucleotides of SEQ ID NO:179. The specification provides guidance and examples for a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NO:179 which encodes a phosphoenolpyruvate carboxykinase from *Corynebacterium glutamicum* consisting of the amino acid sequence of SEQ ID NO:180. The specification discloses that SEQ ID NO:179 consists of 1953 nucleotides and SEQ ID NO:180 consists of 610 amino acid residues. However, the specification does not provide guidance, prediction and working examples for making any nucleic acid molecule comprising a nucleotide sequence which is at least 50% identical to SEQ ID NO:179 or a complement thereof and any nucleic acid molecule which encodes any naturally occurring allelic variant of a polypeptide comprising SEQ ID NO:180...[and] any isolated nucleic [acid] molecule comprising a fragment of at least 15 contiguous nucleotides of SEQ ID NO:179. (Office Action, page 6)

Applicants respectfully traverse the foregoing rejection. As an initial matter, Applicants note that in order to expedite examination, but in no way acquiescing to the validity of the Examiner's rejection, originally numbered claim 39(c) (directed to nucleic acid molecules encoding allelic variants of a polypeptide comprising the amino acid sequence of SEQ ID NO: 180) has been cancelled, thereby rendering the rejection of claim 39, as it relates to allelic variants, moot. Each remaining aspect of the Examiner's rejection is addressed below.

Rejection of Claims Directed to Sequences of 50% Identity to the Nucleotide Sequence of SEQ ID NO:179

With respect to newly numbered claim 39(c) and new claim 60, as they relate to nucleotide sequences of at least 90% or 95% identity to the nucleotide sequence of SEQ ID NO:180, and claims depending therefrom, Applicants respectfully traverse the foregoing rejection for the following reasons. Once again, Applicants direct the Examiner's attention to Example 14 of the Written Description Guidelines, which states that claims directed to sequences of 95% identity to a disclosed sequence and characterized by a particular function are sufficiently described and enabled in accordance with 35 U.S.C. § 112, first paragraph, where the specification discloses an assay for identifying such sequences. Indeed, the present specification provides extensive guidance for making and identifying such sequences. Applicants direct the Examiner's attention to Examples 4-9 at page 51, line 29 to page 57, line 24 of the specification and Example 11 at page 58, line 24 to page 60, line 6 of the specification. Specifically, at page 51, line 29 to page 53, line 16, the specification teaches methods for the *in vivo* mutagenesis of bacterial strains and methods of transferring mutated nucleic acid molecules (*e.g.*, nucleic acid molecules of 90% or 95% identity to the nucleotide sequence of SEQ ID NO:179) into such bacterial strains. In addition, at page 53, line 18 to page 54, line 4, the specification teaches assays for assessing the expression of the mutated protein in the bacterial strains. At page 54, line 6 to page 56, line 28, the specification teaches assays for assessing the activity of the mutated protein (*e.g.*, assays for determining whether the mutated protein retains the activity of the amino acid sequences set forth in SEQ ID NO:180). Furthermore, at page 56, line 30 to page 57, line 24, the specification teaches methods for determining the effect of the mutated protein on the production of a desired product, for example, lysine, from cultured bacteria. Additionally, at page 58, line 24 to page 60, line 6, the specification teaches techniques for identifying sequence identity, for example, for identifying sequences of at least 90% or 95% identity to SEQ ID NO:179. In view of the foregoing, it is evident that the specification provides extensive teachings that would enable one skilled in the art to design and assess the activity of sequences of 90% or 95% identity to the nucleotide sequence of SEQ ID NO:179.

Applicants submit that, even though Example 14 is part of the *Written Description Guidelines* and not the *Enablement Guidelines*, this example does state explicitly that ***one skilled in the art would be able to generate a nucleotide sequence of 95% identity to another nucleotide sequence using only routine experimentation***. Specifically, the relevant section of

Example 14 provides that "*[t]he procedures for making variants of SEQ ID NO:3* ***are conventional in the art*** *and an assay is described which will identify other proteins having the claimed catalytic activity. Moreover, procedures for making variants of SEQ ID NO:3 which have 95% identity to SEQ ID NO:3 and retain its activity are* ***conventional in the art***" **(Emphasis added).**

Accordingly, while the *Written Description Guidelines* are generally directed to describing the standard for satisfying the written description requirement, in this particular example, the Guidelines clearly provide guidance on the USPTO's position regarding a key question for determining whether the enablement requirement has been satisfied: would it be routine for one of skill in the art to generate a sequence with 90% or 95% identity to a specified nucleotide sequence and which retains the activity of that specified nucleotide sequence? The answer to that question, as provided by the USPTO, is: yes. The Guidelines provide that claims to sequences of 95% identity with a functional limitation are sufficiently enabled where the specification provides assays for the identification of such sequences having the requisite function. Accordingly, because it is conventional, *i.e.*, routine, to make sequences of at least 90% or 95% identity and because the instant specification provides assays for identifying nucleic acid sequences of SEQ ID NO:179 that encode a polypeptide having phosphoenolpyruvate carboxykinase activity (see, for example, Examples 4-9 at page 51, line 29 to page 57, line 24 of the specification and Example 11 at page 58, line 24 to page 60, line 6 of the specification), one of skill in the art would be able to make and use the claimed invention using only routine experimentation.

Rejection of Claims Directed to Nucleotide Sequences of 15 Contiguous Nucleotides

With regard to newly numbered claim 39(d) directed to nucleic acid molecules comprising a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179, or the complement thereof, the Examiner is of the opinion that

> experimentation involves selecting any 15 contiguous nucleotides of SEQ ID NO:179 and determining the biological function and use of the selected 15 contiguous nucleotides of SEQ ID NO:179. General teaching regarding screening and searching for the claimed invention using phosphoenolpyruvate carboxykinase assays taught in the specification is not guidance for making the claimed invention. (Office Action, page 6)

Applicants respectfully traverse the foregoing rejection for the following reasons. Initially, Applicants note that, as set forth in M.P.E.P. § 2164.02, the absence of a working

example is not sufficient to undermine the enablement of a claimed invention. Indeed, "the specification need not contain an example if the invention is otherwise disclosed in such manner that one skilled in the art will be able to practice it without an undue amount of experimentation. *In re Borkowski*, 422 F.2d 904." M.P.E.P. § 2164.01(b) further sets forth that "as long as the specification discloses at least one method for making and using the claimed invention that bears a reasonable correlation to the entire scope of the claim, then the enablement requirement of 35 U.S.C. 112 is satisfied. In re Fisher, 427 F.2d 833, 839."

In view of the foregoing enablement standards, Applicants respectfully submit that the teachings of the present specification are sufficient to enable an ordinarily skilled artisan to make and use the claimed invention using only routine experimentation. Indeed, Applicants submit that the present specification contains ample guidance on how to make and use nucleic acid molecules comprising at least 25 contiguous nucleotides of SEQ ID NO:179. At, for example, page 25, line 25 to page 26, line 9 of the specification, Applicants disclose various length fragments of the claimed nucleic acid molecules, as well as their use as probes or primers. In addition, at page 31, lines 16 to page 33, line 9 of the specification, Applicants disclose various length fragments and their use as antisense molecules to, for example, repress translation of RNA. In addition, at page 46, line 5 to page 48, line 24, Applicants disclose use of such probes and primers in, for example, modulating cellular production of a desired compound such as an amino acid, for example, lysine, identifying *Corynebacterium glutamicum* and related organisms, mapping of genomes of organisms related to *Corynebacterium glutamicum*, identifying and localizing *Corynebacterium glutamicum* sequences of interest, evolutionary studies and modulating metabolic pathways. As noted earlier, such fragments do not necessarily need to exhibit an enzymatic activity of the full length sequence. Indeed, probes, primers and antisense molecules do not need to exhibit phosphoenolpyruvate carboxykinase activity. Accordingly, assessing biological function, as the Examiner asserts is required, is unnecessary.

In addition, Applicants respectfully submit that the present specification teaches techniques for the generation of such fragments, for example, using standard synthetic techniques such as an automated DNA synthesizer (see, for example, page 23, lines 12-37 of the specification) or using chemical synthesis and enzymatic ligation reactions (see, for example, page 31, lines 24-26). Moreover, standard techniques well known by a skilled artisan, for example, the use of restriction enzymes to cleave at specific sites of a DNA sequence, could also be used.

With respect to fragments retaining a biological function, Applicants submit that the teachings of the present specification, alone or in combination with the state of the art at the time of the filing of the present application, are sufficient to allow a skilled artisan to make and use such biologically active fragments. Specifically, the specification, for example, at page 26, line 33 to page 27, line 10, teaches techniques for preparing nucleic acid fragments that encode biologically active portions of the polypeptides of the invention by incorporating an active domain known to be important for phosphoenolpyruvate carboxykinase activity. For example, as set forth above, the domain [F/Y]-P-S-[A/G/M/S]-C-G-K-T-[N/S] (amino acid residues 270-278 of SEQ ID NO:2 encoded by nucleotide residues 908-934 of SEQ ID NO:1) had been identified at the time of filing of the present application as a highly conserved phosphoenolpyruvate carboxykinase domain that is involved in enzymatic activity. Moreover, the specification, for example, at Example 8, at page 56, lines 1-28 of the specification, in addition to standard techniques known in the art at the time of filing of the present application, provide assays for confirming that such fragments of at least 25 contiguous nucleotides retain a phosphoenolpyruvate carboxykinase activity.

For each of the foregoing reasons, Applicants respectfully submit that one skilled in the art would be able to make and use the nucleic acid molecules of the present invention using only routine experimentation. Accordingly, Applicants respectfully request reconsideration and withdrawal of the rejection of the pending claims under 35 U.S.C. § 112, first paragraph.

Rejection of Claim 43 Under 35 U.S.C. § 112, First Paragraph

The Examiner has rejected claim 43 under 35 U.S.C. §112 as "failing to comply with the enablement requirement." In particular, the Examiner is of the opinion that

> [t]he nature and breadth of the claims encompass transgenic plants and animals including humans transformed with the claimed vector, where the vector comprises the nucleic acid molecule of claim 39.
>
> While the specification provides guidance for transforming isolated *E.coli* host cells with the vector, the specification does not provide guidance, prediction, and working examples for making transgenic plants and animals including humans transformed with the claimed vector. Thus, an undue amount of trial and error experimentation must be performed to make the claimed transgenic plants and animals including humans and determining whether the claimed vector expresses the claimed nucleic acid molecules. (Office Action, page 7)

Applicants respectfully traverse the foregoing rejection for the following reasons. Applicants submit that the specification provides sufficient teachings to enable one of skill in the art to make and use the claimed invention using only routine experimentation. Specifically, Applicants submit that the specification teaches and enables the production and use of a plethora of host cells, including methods of making and using transgenic plants and animals containing the vectors of the present invention. For example, at page 33, line 17 to page 40, line 24 of the specification, Applicants provide extensive teachings on the transfection of various host cells including, for example, insect cells, unicellular plant cells such as algae, multicellular plant cells, mammalian cells and fungal cells such as yeast. Moreover, at Table 3 of the specification, Applicants teach a vast number of bacterial strains, including, for example, those from the genus *Brevibacterium* and *Corynebacterium* which could be used in the claimed methods. Furthermore, the specification provides extensive teachings on the selection and design of vectors based on the particular host cell to be utilized (see page 33, line 17 to page 40, line 24 of the specification). Indeed, the specification teaches the selection and design of expression vectors, including, for example, regulatory elements such as promoters, enhancers and the like, for transforming prokaryotes such as bacterial cells (at page 35, line 8 to page 36, line 19 of the specification), yeast cells (at page 36, lines 20-30 of the specification), insect cells (at page 36, lines 31-35 of the specification), unicellular and multicellular plant cells (at page 36, line 36 to page 37, line 7 of the specification) and mammalian cells (at page 37, line 8 to page 37, line 34 of the specification). The specification further provides teachings on techniques for the introduction of such vectors into either prokaryotic or eukaryotic cells, for example, by transformation, transfection, conjugation and transduction (see page 38, line 27 to page 39, line 2 of the specification) and techniques for identifying and selecting those cells that are successfully transfected, for example, by use of a selectable marker (see page 39, lines 3-13 of the specification). In addition, Examples 4 and 5, at pages 51, line 29 to page 53, line 14 of the specification, as described above, further elaborate on methods of transfecting host cells with the modified nucleic acid sequences of the invention. In view of the foregoing extensive teachings in Applicants' specification as well as the general knowledge in the art at the time of the invention, one skilled in the art would be able to make and use the claimed invention using only routine experimentation.

Notwithstanding the foregoing, in order to expedite examination, but in no way acquiescing to the validity of the Examiner's rejection, Applicants have amended claim 43 to be

directed to host cells which are microorganisms, the subject matter of claim 44 which the Examiner indicates is enabled. Applicants submit that the foregoing amendment renders the rejection of claim 43, and its corresponding dependent claims, moot. Accordingly, Applicants respectfully request reconsideration and withdrawal of the rejection of claim 43 under 35 U.S.C. § 112, first paragraph.

Rejection of Claim 39 Under 35 U.S.C. § 102(e)

The Examiner has rejected claim 39 under 35 U.S.C. §102(e) as being anticipated by Fleischman *et al.* (USPN 6,294,328) (hereinafter referred to as "Fleischman"). In particular, the Examiner is of the opinion that

> Fleischman et al. (US Patent 6,294,328) teach an isolated nucleic acid molecule encoding an amino acid sequence that is 65% identical to SEQ ID NO:180...
>
> Because claim 1 [sic] does not specifically state a biological function and specifically state that the encoded naturally occurring allelic variant must have 100% amino acid identity to SEQ ID NO: 180, than [sic] the examiner takes the position that the teachings of Fleischman et al. anticipate the claim. (Office Action, page 10)

Applicants traverse the foregoing rejection for the following reasons. For a prior art reference to anticipate, in terms of 35 U.S.C. § 102, a claimed invention, the prior art must teach each and every element of the claimed invention. *Lewmar Marine v. Barient*, 827 F.2d 744, 3 USPQ2d 1766 (Fed. Cir. 1987). As amended, claim 39, is directed to (a) an isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:179, or the complement thereof; (b) an isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:180, or the complement thereof; (c) an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 90% identical to the entire nucleotide sequence of SEQ ID NO:179 and wherein the nucleic acid molecule encodes a polypeptide having a phosphoenolpyruvate carboxykinase activity, or the complement thereof; and (d) an isolated nucleic acid molecule comprising a fragment of at least 25 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179, or the complement thereof.

Applicants respectfully submit that Fleischman fails to teach or suggest each and every element of the claimed invention. With regard specifically to the Examiner's rejection of claim 39 as it relates to allelic variants, Applicants submit that Fleischman fails to characterize or identify features of the disclosed nucleotide sequence in any manner that would indicate that the

sequence disclosed therein encodes an allelic variant of SEQ ID NO:180, as set forth in the present application. Indeed, an allelic variant is defined as "produced by or occurring in alleles of the same gene; applied to different forms of a protein, to mutations, etc." (see *The Encyclopedia of Molecular Biology* (1994), enclosed herein as Appendix A) and is generally known in the art as a sequence which varies by 1-5% on the nucleic acid level yet possesses a similar function, as set forth above. Fleischmann fails to provide any characterization of the disclosed nucleotide sequence. In fact, Fleischmann fails to even identify the open reading frames of the sequence disclosed as SEQ ID NO:2. Indeed, Fleischmann merely presents the complete DNA sequence of the CDC 1551 strain of *M. tuberculosis* (SEQ ID NO:2), without any characterization of open reading frames existing within the sequence. Indeed, there is no indication that the nucleotide sequence relied upon by the Examiner is even translated, let alone into an allelic variant of a polypeptide comprising SEQ ID NO:180.

Notwithstanding the foregoing, in order to expedite examination, but in no way acquiescing to the validity of the Examiner's rejection, Applicants have cancelled originally numbered claim 39(c), which relates to allelic variants of polypeptides comprising the amino acid sequence of SEQ ID NO:180, thereby rendering the rejection of claim 39(c), and its corresponding dependent claims, moot.

With respect to the remaining claims, as indicated in the attached local and global alignments of nucleotide residues 251891-253705 of SEQ ID NO:2 of Fleischmann with SEQ ID NO:179 of the present application (attached herein as Appendices J and K, respectively), Fleischmann fails to disclose a nucleotide sequence of at least 25 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179 or a nucleotide sequence of at least 90% or 95% identity to the entire nucleotide sequence of SEQ ID NO:179 that encodes a polypeptide having a phosphoenolpyruvate carboxykinase activity, as required by the pending claims.

Accordingly, Applicants respectfully request reconsideration and withdrawal of the rejection of the pending claims under 35 U.S.C. § 102(e).

Rejection of Claims 39, 41-44 and 46 Under 35 U.S.C. § 102(b)

The Examiner has also rejected claims 39, 41-44 and 46 under 35 U.S.C. §102(b) as being anticipated by Bauer *et al.* (USPN 5,604,116) (hereinafter referred to as "Bauer"). In particular, the Examiner is of the opinion that

> [Bauer] et al. teach an isolated nucleic acid molecule that comprises 18 consecutive nucleotides of SEQ ID NO:179..., vectors and host cells comprising said isolated nucleic acid molecule, and a method for expressing said nucleic acid molecule in a host cell such as *E.coli* to produce a polypeptide... Thus, the reference teachings anticipate the claimed invention. (Office Action, page 10)

Applicants traverse the foregoing rejection for the following reasons. For a prior art reference to anticipate, in terms of 35 U.S.C. § 102, a claimed invention, the prior art must teach each and every element of the claimed invention. *Lewmar Marine v. Barient*, 827 F.2d 744, 3 USPQ2d 1766 (Fed. Cir. 1987). As indicated above, claim 39, as amended, is directed to (a) an isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:179, or the complement thereof; (b) an isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:180, or the complement thereof; (c) an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 90% identical to the entire nucleotide sequence of SEQ ID NO:179 and wherein the nucleic acid molecule encodes a polypeptide having a phosphoenolpyruvate carboxykinase activity, or the complement thereof; and (d) an isolated nucleic acid molecule comprising a fragment of at least 25 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179, or the complement thereof.

Applicants respectfully submit that Bauer fails to teach or suggest each and every element of the claimed invention. Specifically, Applicants respectfully submit that Bauer does not disclose a nucleotide sequence of at least 25 contiguous nucleotides of SEQ ID NO:179 as asserted by the Examiner. In support, Applicants submit the following alignments:

a) local and global alignments of SEQ ID NO:179 of the present application versus SEQ ID NO: 398 of Bauer (attached herein as Appendices B and C, respectively); and

b) local and global alignments of SEQ ID NO:179 of the present application versus the reverse complement of SEQ ID NO: 398 of Bauer (attached herein as Appendices D and E, respectively).

In addition, Applicants submit that the sequence of 18 contiguous nucleotides asserted to be shared by the Examiner, as indicated in the alignment forwarded by the Examiner, does not exist in either the nucleotide sequence of SEQ ID NO:179 or the reverse complement thereof. In support, Applicants submit the following alignments:

a) local and global alignments of alleged shared sequence of 18 nucleotides versus SEQ ID NO:179 (attached herein as Appendices F and G); and

b) local and global alignments of alleged shared sequence of 18 nucleotides versus the reverse complement of SEQ ID NO:179 (attached herein as Appendices H and I).

As demonstrated by the foregoing alignments, Bauer fails to teach a nucleic acid molecule comprising a fragment of at least 25 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:179. Accordingly, Applicants respectfully request reconsideration and withdrawal of the rejection of the pending claims under 35 U.S.C. § 102(b).

CONCLUSION

In view of the foregoing remarks, reconsideration of the rejections and allowance of all pending claims is respectfully requested. If there are any remaining issues or if the Examiner believes that a telephone conversation with Applicants' Attorney would be helpful in expediting prosecution of this application, the Examiner is invited to call the undersigned at (617) 227-7400.

The Commissioner is hereby authorized to charge any deficiency in the fees paid herewith, or credit any overpayment, to Deposit Account No. 12-0080, under Order No. BGI-126CPCN, from which the undersigned is authorized to withdraw.

Dated: **June 20, 2006**

Respectfully submitted,

By _____

Maria Laccotripe Zacharakis, Ph.D., J.D.
Registration No.: 56,266
LAHIVE & COCKFIELD, LLP
28 State Street
Boston, Massachusetts 02109
(617) 227-7400
(617) 742-4214 (Fax)
Attorney/Agent For Applicants

THE ENCYCLOPEDIA OF Molecular Biology

EDITOR IN CHIEF
SIR JOHN KENDREW

EXECUTIVE EDITOR
ELEANOR LAWRENCE

Blackwell Science

Blackwell Science Ltd
Editorial Offices:
Osney Mead, Oxford OX2 0EL
25 John Street, London WC1N 2BL
23 Ainslie Place, Edinburgh EH3 6AJ
238 Main Street, Cambridge,
Massachusetts 02142, USA
54 University Street, Carlton,
Victoria 3053, Australia

Other Editorial Offices:
Arnette Blackwell SA
1, rue de Lille, 75007 Paris
France

Blackwell Wissenschafts-Verlag GmbH
Kurfürstendamm 57
10707 Berlin, Germany

Feldgasse 13, A-1238 Wien
Austria

First published 1994
Reprinted as a paperback 1995

Set by Semantic Graphics, Singapore
Printed in Great Britain at the Alden Press Limited,
Oxford and Northampton
and bound by Hartnolls Ltd, Bodmin, Cornwall

DISTRIBUTORS

Marston Book Services Ltd
PO Box 87
Oxford OX2 0DT
(*Orders*: Tel: 01865 791155
Fax: 01865 791927
Telex: 837515)

USA
Blackwell Science, Inc.
238 Main Street
Cambridge, MA 02142
(*Orders*: Tel: 800 215-1000
617 876-7000
Fax: 617 492-5263)

Canada
Oxford University Press
70 Wynford Drive
Don Mills
Ontario M3C 1J9
(*Orders*: Tel: 416 441-2941)

Australia
Blackwell Science Pty Ltd
54 University Street
Carlton, Victoria 3053
(*Orders*: Tel: 03 347-0300
Fax: 03 349-3016)

A catalogue record for this title
is available from both the British Library
and the Library of Congress

ISBN 0-632-02182-9
0-86542-621-X (pbk)

The term indel is a neutral expression which avoids indicating the direction of change. However, as shown below, it is possible to just keep adding gaps to the first sequence, thus increasing the proportion of matches for this sequence to 7/8 for the first part of sequence A.

```
Sequence A   T-G----AA-C--------TC---C-ACCAAAAAGGAAGACT
Sequence B   TAGGCTCACGCACCAACATCATACT
             * *    * *        **    * 7/8
```

To avoid the absurdity of any sequence showing a high similarity to every other sequence a gap penalty is introduced and is subtracted from the number of matches. It is often a two-parameter penalty where a gap spans more than one nucleotide; there is a higher weighting for introducing a gap and a lesser penalty for extending a gap. As yet, the weights given the gap penalty are arbitrary in that values are estimated, rather than calculated from first principles. It is probably possible to calculate the expected value for a gap penalty but impractical in practice, so values are usually estimated from simulations. In this approach, many pairs of random sequences could be generated. For each pair of random sequences the maximum number of matches would be found and the increase in number of matches counted as gaps (indels) are added.

The calculation becomes even more complicated for 'multiple alignment' where three or more sequences are compared. The optimum alignment is dependent on the tree which relates the sequences to each other because it improves the score if a single indel can account for changes to two or more sequences.

Other information useful for sequence alignment includes finding similar two-dimensional folding, particularly in RNA sequences; three-dimensional similarities, particularly in proteins; and similar positions of intron/exon boundaries. Alignment remains a major computational problem but is a vital part of a molecular study. Conserved regions are, as predicted by the neutral theory (*see* MOLECULAR EVOLUTION), the important functional parts of macromolecules so sequence alignment (the phylogenetic method) is an important aspect of any study of a new gene sequence.

Waterman, M.S. et al. (1991) In *Phylogenetic Analysis of DNA Sequences* (Miyamoto, M.M. & Cracraft, J., Eds) 59–72 (Oxford University Press, Oxford).

alkaline phosphatase Hydrolytic enzyme (EC 3.1.3.1) used to catalyse the removal of 5′-phosphate residues from nucleic acids. It is a dimeric GLYCOPROTEIN consisting of two identical subunits (subunit M_r 69 000) and four atoms of zinc per molecule. Usually obtained from either *Escherichia coli* (BAP) or calf intestine (CIP), it has two major applications in molecular biology:

1. To remove 5′-phosphate residues from a linearized double-stranded DNA to prevent subsequent self-ligation.
2. To remove the 5′-phosphate residue from a single-stranded DNA or RNA molecule to allow subsequent labelling of the 5′-end using T4 polynucleotide kinase (*see* END-LABELLING).

It is also used as a component of a conjugated antibody immunodetection system.

alkaptonuria An AUTOSOMAL RECESSIVE inborn error of metabolism resulting from a deficiency of homogentisic acid oxidase which catalyses the oxidation of homogentisic acid to maleylacetoacetic acid. The homogentisic acid slowly polymerizes and the resulting brown-black product may be seen deposited in cartilage, tendons, and ligaments. Ochronotic arthritis may develop. The GENE FREQUENCY is 1 : 1000.

alkylating agent Any of a large group of mutagens and carcinogens which generally act by covalent modification of DNA to generate cytotoxic and mutagenic adducts. Their action typically involves methylation or ethylation of nitrogen and oxygen atoms in the polynucleotide. Some environmental mutagens such as nitrosamines require metabolic activation by mixed function oxidases (CYTOCHROME P450) to be converted into reactive species. Certain anticancer drugs such as cyclophosphamide become alkylating aents by activation in this way. Other agents in clinical use (nitrosoureas, nitrogen mustards) are used directly as reactive compounds. The short-lived alkylating intermediate itself arises when an alkylating agent decomposes spontaneously in aqueous solution to produce an alkyl carbonium cation that reacts with DNA. *See*: CHEMICAL CARCINOGENS AND CARCINOGENESIS: DNA REPAIR.

ALL ACUTE LYMPHOBLASTIC LEUKAEMIA.

allele A variant form of a given GENE. The term was first coined by Gregor Mendel to describe the alternative forms of the inheritable 'factors' that encoded a particular characteristic of the pea plant. For example, yellow and green peas carry different alleles of the gene determining pea colour (*see* MENDELIAN INHERITANCE). A large number of genes are present in two or more allelic forms in the population and some genes, notably those encoding MHC MOLECULES, have numerous alleles. A diploid organism carrying two different alleles of a gene is said to be heterozygous for that gene, whereas a homozygote carries two copies of the same allele.

The geneticist H.J. Muller coined the terms neomorphic allele (neomorph), hypomorphic allele (hypomorph), hypermorphic allele (hypermorph), amorphic allele (amorph), and antimorphic allele (antimorph), to classify different mutations of a given gene. The effects of the test mutation when homozygous, and when heterozygous with a complete deletion of the gene, were compared with the effects of a homozygous deletion of the gene: that is, *mutation/mutation* is compared with *mutation*/deletion and with deletion/deletion. The results of the combinations are interpreted as:

1 Amorphic allele: when *mutation*/deletion shows an effect identical to a homozygous deletion of a gene. This indicates that the mutant allele has no gene activity.
2 Antimorphic allele: when *mutation/mutation* has a more severe phenotype than *mutation*/deletion. This indicates that the mutant allele possesses activity opposite to that of wild type.
3 Hypermorphic allele: these alleles possess a greater than wild-type activity, but are difficult to detect because most are fully recessive and would produce a phenotype when homozygous which would be indistinguishable from wild type.

4 Hypomorphic allele: when *mutation*/deletion has a phenotype which is similar to but more extreme than *mutation/mutation*. This indicates an allele with reduced activity compared with wild type. 5 Neomorphic allele: possesses qualitatively different activity from wild type. Such alleles are almost always dominant to both wild-type and deficiency alleles.

allelic Produced by or occurring in alleles of the same gene; applied to different forms of a protein, to mutations, etc.

allelic exclusion The situation in which only one of the pair of ALLELES at a genetic LOCUS is expressed in a diploid cell. It is characteristic of the expression of the immunoglobulin loci and T cell receptor loci in B CELLS and T CELLS respectively (*see* B CELL DEVELOPMENT; IMMUNOGLOBULIN GENES; T CELL DEVELOPMENT; T CELL RECEPTOR GENES). In B cells, for example, there will be a productive rearrangement of the V_H, D, and J_H gene segments leading to the synthesis of a heavy chain polypeptide at only one of the two immunoglobulin heavy chain alleles (*see* V-(D)-J RECOMBINATION). The heavy chain allele on the other chromosome of the homologous pair will either have its V_H, D, and J_H segments in the unrearranged germ-line configuration, or will carry some form of incomplete or otherwise abortive rearrangement such that this second allele does not yield a heavy chain polypeptide that can be expressed on the cell surface. The choice of parental allele for expression is apparently random. Allelic exclusion occurs in a similar fashion at the immunoglobulin light chain loci and at the T cell receptor loci, and ensures that each B and T lymphocyte expresses antigen receptors of a single antigen specificity. *See also*: ISOTYPIC EXCLUSION.

allelic rescue A method for directly cloning a mutant ALLELE of a given gene from the genome of SACCHAROMYCES CEREVISIAE. It involves TRANSFORMATION of *S. cerevisiae* with a plasmid carrying a wild-type allele of the target gene linearized by deletion of a DNA sequence internal to the target gene. Upon transformation into a strain of *S. cerevisiae* carrying the desired mutant allele, the gap in the transforming plasmid is spontaneously repaired by the host cell's DNA synthesis machinery using the chromosomally located mutant gene as a template (gap repair). The plasmid now contains a copy of the mutant allele and can be recovered from the culture.

Orr-Weaver, T.L. et al. (1982) *Methods Enzymol.* 101, 228–245.

alloantigens Self molecules that may be recognized as foreign between members of the same species. Such molecules usually have the same or similar function but are products of different ALLELES of the same gene, for example, T CELL RECEPTORS recognizing ALLO-MHC molecules. Here the alloantigen is in many cases specified by a particular peptide bound in the peptide-binding site.

alloantisera ANTIBODIES raised in one member of a species that specifically recognize molecules of a genetically distinct member of the same species. Human alloantisera are produced by multiparous women who have been exposed to paternal allotypes from the foetus, multiply transfused individuals and immunized volunteers.

alloenzymes Variant forms of an enzyme found in different individuals of the same species and resulting from the existence of multiple alleles within the population.

allogeneic, allogenic Describes genetic differences within species, that is, differences between individuals of the same species. Allogeneic cells are therefore simply those from another individual of the same species, but the term is used particularly in immunology to describe differences in MHC MOLECULES (*see* MAJOR HISTOCOMPATIBILITY COMPLEX) on lymphocytes from genetically nonidentical individuals (*see* ALLO-MHC). Allogeneic cells have been used to investigate mechanisms of tolerance and the specificity of the immune response. Thus, injecting semi-allogeneic parental cells into an F1 hybrid mouse might be expected to cause a one-way GRAFT-VERSUS-HOST response.

allograft A graft of tissue between two genetically nonidentical individuals of the same species. Allografts are generally rejected in vertebrates because of the immune response mounted against the foreign tissue (*see*: ALLO-MHC; MAJOR HISTOCOMPATIBILITY COMPLEX), but are generally accepted in invertebrates.

allo-MHC Allogeneic MHC, which describes a situation of genetic nonidentity at the loci of the MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) between individuals of the same species. Humans, other than identical twins, are related allogeneically, as are outbred mice and mice from different inbred strains. Allogeneic individuals display different allelic forms of at least some MHC antigens (MHC MOLECULES) on their cell surfaces, so that tissue grafts exchanged between such individuals are rejected. Immune cells from different individuals also cannot normally cooperate to produce an MHC-restricted immune response.

allophenic mice CHIMAERIC or MOSAIC mice generated by aggregating two or more genetically distinct cleavage-stage mouse embryos and allowing them to develop to term in a foster mother. This experimental system was developed in the 1960s for investigating CELL LINEAGE, cell allocation, cell fusions and CELL–CELL INTERACTIONS during development.

allophycocyanin PHYCOBILIPROTEIN pigment that forms the core of the PHYCOBILISOMES of red algae and CYANOBACTERIA. The chromophore is phycocyanobilin, an open-chain tetrapyrrole, which is covalently linked to the apoprotein through thioether linkages. Like other phycobiliproteins, the protein is composed of an $\alpha\beta$ monomer which forms trimeric and hexameric aggregates. Its absorption maximum at 650 nm (670 nm in allophycocyanin B) lies at lower energy (longer wavelength) than those of PHYCOERYTHRIN and PHYCOCYANIN, so facilitating excitation energy transfer from the latter pigments to allophycocyanin and hence, via the linker protein, to the reaction centre of PHOTOSYSTEM II. *See also*: PHOTOSYNTHESIS.

alloreactivity The immune reactivity shown by an individual to

APPENDIX B

lalign output for SEQ ID NO:179 vs. SEQ ID NO:389 (USPN 5604116)

[ISREC-Server] Date: Mon Jun 19 19:33:46 Europe/Zurich 2006

LALIGN finds the best local alignments between two sequences version 2.0u66 September 1998 Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381 resetting to DNA matrix

```
 Comparison of:
(A) ./wwwtmp/lalign/.20719.1.seq SEQ ID NO:179
(B) ./wwwtmp/lalign/.20719.2.seq SEQ ID NO:389 (USPN 5604116)
 using matrix file: DNA, gap penalties: -14/-4

  63.6% identity in 55 nt overlap; score:   75 E(10,000): 5.5e+02

    1330      1340      1350      1360      1370      1380
SEQ    CAAGATCGACGCAATCCTCTTCGGTGGACGTCGCGCAGACACCG-TCCCACTGGT
       :: :::::: : :::: ::  :     :: :   :::: ::::: : :: ::: :
SEQ    CATGATCGATGAAATCATCACCC----ACCTGAAGCAGCCACCGCTGCCGCTGCT
        20        30        40             50        60
```

```
  68.4% identity in 38 nt overlap; score:   72 E(10,000): 9.6e+02

        410       420       430       440
SEQ    CTGGGCTCCACCACAGGCAATGAAGGACGAAATGTCCA
       :::: :: :: ::    :::::: : : :: ::::: :
SEQ    CTGGACTTCAACAACCTCAATGACG-AAGACATGTCTA
         70        80        90       100
```

```
  57.7% identity in 168 nt overlap; score:   71 E(10,000): 1.2e+03

         1680      1690          1700      1710      1720      1730
SEQ    TCATCGACCGCATCGAAG--GCCAC-GTTGGCGCAGACGAGACCGTTGTTGGACACACCG
       :::::  :: : : :::: :::::: : :: ::: : :  :::     ::  ::: :::
SEQ    TCATCACCCACCT-GAAGCAGCCACCGCTGCCGCTG-CTGGAC-----TTCAACA-ACCT
              40          50        60         70                80

             1740        1750      1760              1770
SEQ    CTAAGGCCGAAGACCT--CGACCTCGACGGCC--------TCGACACCCCAA--TTGAGG
       : :: : :::::::: :  : : :  :: :: :         ::: :  :: ::  : ::::
SEQ    C-AATGACGAAGACATGTCTATCCTGATGGACAATAACCTTCGTCGTCCAAACCTCGAGG
              90       100       110       120       130       140

     1780      1790      1800         1810      1820
SEQ    ATGTCAAGGAAGCACTGACCGCTCCTGCAGA--GCAGTGGGCAAACGA
          ::::    ::  : :   ::: ::::::  ::: :  :::: ::
SEQ    CATTCAACCGTGCTGTCAAGTCTC-TGCAGAATGCA-TCAGCAATTGA
            150       160       170       180
```

Back to ISREC bioinformatics group home page

APPENDIX C

lalign output for SEQ ID NO:179 vs. SEQ ID NO:389 (USPN 5604116)

[ISREC-Server] Date: Mon Jun 19 19:33:58 Europe/Zurich 2006

---

resetting matrix to DNA ./wwwtmp/lalign/.20719.1.seq : 1953 nt

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
SEQ ID NO:179                                        1953 nt vs.
SEQ ID NO:389 (USPN 5604116)                          339 nt
scoring matrix: DNA, gap penalties: -14/-4
15.0% identity;         Global alignment score: -5745

               10         20         30         40         50         60
./wwwt ATGTGTCCGTTGTCTCACCTAAAGTTTTAACTAGTTCTGTATCTGAAAGCTACGCTAGGG
       :::                                                 ::::
SEQ    ATG-------------------------------------------------GCTA--------


               70         80         90        100        110        120
./wwwt GGCGAGAACTCTGTCGAATGACACAAAATCTGGAGAAGTAATGACTACTGCTGCAATCAG
                                                          ::::::
SEQ    ------------------------------------------------ACTGCT--------
                                                           10

              130        140        150        160        170        180
./wwwt GGGCCTTCAGGGCGAGGCGCCGACCAAGAATAAGGAACTGCTGAACTGGATCGCAGACGC
          ::                                          :::  :::::
SEQ    ----CT--------------------------------------AACATGATCG-------
                                                       20

              190        200        210        220        230        240
./wwwt CGTCGAGCTCTTCCAGCCTGAGGCTGTTGTGTTCGTTGATGGATCCCAGGCTGAGTGGGA
                                             ::: :                 :
SEQ    ---------------------------------------ATGAA----------------A
                                                30

              250        260        270        280        290        300
./wwwt TCGCATGGCGGAGGATCTTGTTGAAGCCGGTACCCTCATCAAGCTCAACGAGGAAAAGCG
       ::             :::               :::: : : ::::             :::
SEQ    TC------------ATC--------------ACCCACCTGAAGC------------AGCC
                                      40                        50

              310        320        330        340        350        360
./wwwt TCCGAACAGCTACCTAGCTCGTTCCAACCCATCTGACGTTGCGCGCGTTGAGTCCCGCAC
        :::                              ::: :: :::     : ::
SEQ    ACCG----------------------------CTGCCGCTGC-----TGGA---------
                                        60             70

              370        380        390        400        410        420
./wwwt CTTCATCTGCTCCGAGAAGGAAGAAGATGCTGGCCCAACCAACAACTGGGCTCCACCACA
       :::::                                       :::::    :::
SEQ    CTTCA-------------------------------------ACAAC----CTC-------
```

```
                                                  80

              430       440       450       460       470       480
./wwwt GGCAATGAAGGACGAAATGTCCAAGCATTACGCTGGTTCCATGAAGGGGCGCACCATGTA
              :: ::::::
SEQ    -------AATGACGAA-------------------------------------------
                  90

              490       500       510       520       530       540
./wwwt CGTCGTGCCTTTCTGCATGGGTCCAATCAGCGATCCGGACCCTAAGCTTGGTGTGCAGCT

SEQ    ------------------------------------------------------------


              550       560       570       580       590       600
./wwwt CACTGACTCCGAGTACGTTGTCATGTCCATGCGCATCATGACCCGCATGGGTATTGAAGC
                        : :::::: :: :       :::
SEQ    ------------------GACATGTCTATCC------TGA------------------
                            100

              610       620       630       640       650       660
./wwwt GCTGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTGCCTCCACTCCGTTGGTGCTCC
        ::::::: : :                :::::::       ::::  ::
SEQ    --TGGACAATAAC--------------CTTCGTCG------TCCAAACC-----------
       110       120                           130

              670       680       690       700       710       720
./wwwt TTTGGAGCCAGGCCAGGAAGACGTTGCATGGCCTTGCAACGACACCAAGTACATCACCCA
        : :::                   ::::                          ::: ::
SEQ    --TCGAG-----------------GCAT-----------------------TCAACC-
          140                                               150

              730       740       750       760       770       780
./wwwt GTTCCCAGAGACCAAGGAAATTTGGTCCTACGGTTCCGGCTACGGCGGAAACGCAATCCT

SEQ    ------------------------------------------------------------


              790       800       810       820       830       840
./wwwt GGCAAAGAAGTGCTACGCACTGCGTATCGCATCTGTCATGGCTCGCGAAGAAGGATGGAT
                :::::                     :::: : :::  : ::::
SEQ    ---------GTGCT-------------------GTCAAGTCTCT-GCAGAA--------
                                        160       170

              850       860       870       880       890       900
./wwwt GGCTGAGCACATGCTCATCCTGAAGCTGATCAACCCAGAGGGCAAGGCGTACCACATCGC
                :::               ::::         ::::           : :
SEQ    -----------TGC-------------ATCA---------GCAA-----------TTG-
                                   180

              910       920       930       940       950       960
./wwwt AGCAGCATTCCCATCTGCTTGTGGCAAGACCAACCTCGCCATGATCACTCCAACCATCCC
       :: :::::::
SEQ    AG-AGCATTC-------------------------------------------------
        190

              970       980       990      1000      1010      1020
./wwwt AGGCTGGACCGCTCAGGTTGTTGGCGACGACATCGCTTGGCTGAAGCTGCGCGAGGACGG

SEQ    ------------------------------------------------------------
```

```
             1030      1040      1050      1060      1070      1080
./wwwt CCTCTACGCAGTTAACCCAGAAAATGGTTTCTTCGGTGTTGCTCCAGGCACCAACTACGC
                  ::::      ::::                ::::  :
SEQ    -----------TTAA------AAAT----------------CTCCTG-------------
                   200                          210

             1090      1100      1110      1120      1130      1140
./wwwt ATCCAACCCAATCGCGATGAAGACCATGGAACCAGGCAACACCCTGTTCACCAACGTGGC
                               :::::
SEQ    -----------------------CCATG---------------------------------

             1150      1160      1170      1180      1190      1200
./wwwt ACTCACCGACGACGGCGACATCTGGTGGGAAGGCATGGACGGCGACGCCCCAGCTCACCT
                          ::::                         ::::  :
SEQ    --------------------TCTG----------------------CCCCTG-------
                                                      220

             1210      1220      1230      1240      1250      1260
./wwwt CATTGACTGGATGGGCAACGACTGGACCCCAGAGTCCGACGAAAACGCTGCTCACCCTAA
                    :: ::: : : ::::               ::::
SEQ    -------------GCCACGGCCGCACCC---------------ACGC------------
                    230       240

             1270      1280      1290      1300      1310      1320
./wwwt CTCCCGTTACTGCGTAGCAATCGACCAGTCCCCAGCAGCAGCACCTGAGTTCAACGACTG
                          :::
SEQ    ----------------------GAC-----------------------------------

             1330      1340      1350      1360      1370      1380
./wwwt GGAAGGCGTCAAGATCGACGCAATCCTCTTCGGTGGACGTCGCGCAGACACCGTCCCACT
                   :::    ::::::
SEQ    -------------ATC----CAATCC-----------------------------------
                    250

             1390      1400      1410      1420      1430      1440
./wwwt GGTTACCCAGACCTACGACTGGGAGCACGGCACCATGGTTGGTGCACTGCTCGCATCCGG
                                          ::                  :::
SEQ    ----------------------------------AT-----------------ATC---
                                                            260

             1450      1460      1470      1480      1490      1500
./wwwt TCAGACCGCAGCTTCCGCAGAAGCAAAGGTCGGCACACTCCGCCACGACCCAATGGCAAT
                               :::: :
SEQ    -----------------------AAGGAC-------------------------------

             1510      1520      1530      1540      1550      1560
./wwwt GCTCCCATTCATTGGCTACAACGCTGGTGAATACCTGCAGAACTGGATTGACATGGGTAA
                               :::::             ::::: :::
SEQ    -----------------------GGTGA------------CTGGAATGA---------
                               270                  280

             1570      1580      1590      1600      1610      1620
./wwwt CAAGGGTGGCGACAAGATGCCATCCATCTTCCTGGTCAACTGGTTCCGCCGTGGCGAAGA
                    :: :: ::          :: :::::
SEQ    ---------------ATTCCGTC----------GTAAACTGA-----------------
```

```
                                                290

             1630       1640       1650       1660       1670       1680
./wwwt TGGACGCTTCCTGTGGCCTGGCTTCGGCGACAACTCTCGCGTTCTGAAGTGGGTCATCGA
                                                 : ::::           :::
SEQ    ----------------------------------------CCTTCT----------ATCT-
                                               300

             1690       1700       1710       1720       1730       1740
./wwwt CCGCATCGAAGGCCACGTTGGCGCAGACGAGACCGTTGTTGGACACACCGCTAAGGCCGA
                                                                     ::
SEQ    ------------------------------------------------------------GA
                                                                    310

             1750       1760       1770       1780       1790       1800
./wwwt AGACCTCGACCTCGACGGCCTCGACACCCCAATTGAGGATGTCAAGGAAGCACTGACCGC
       : :::: :
SEQ    AAACCTTG-----------------------------------------------------


             1810       1820       1830       1840       1850       1860
./wwwt TCCTGCAGAGCAGTGGGCAAACGACGTTGAAGACAACGCCGAGTACCTCACTTTCCTCGG
                                      :: :::::
SEQ    -------------------------------GAGAACGC---------------------
                                     320

             1870       1880       1890       1900       1910       1920
./wwwt ACCACGTGTTCCTGCAGAGGTTCACAGCCAGTTCGATGCTCTGAAGGCCCGCATTTCAGC
                                                  : ::::        ::: :
SEQ    ----------------------------------------GCAGGC-------TCAAC
                                                330

             1930       1940       1950
./wwwt AGCTCACGCTTAAAGTTCACGCTTAAGAACTGC
       ::
SEQ    AG-------------------------------
```

Back to ISREC bioinformatics group home page

APPENDIX D

lalign output for SEQ ID NO:179 vs. Reverse Complement of SEQ ID NO:389 (USPN 5604116)

[ISREC-Server] Date: Mon Jun 19 19:33:07 Europe/Zurich 2006

LALIGN finds the best local alignments between two sequences version 2.0u66 September 1998 Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381 resetting to DNA matrix

```
 Comparison of:
(A) ./wwwtmp/lalign/.20685.1.seq SEQ ID NO:179
(B) ./wwwtmp/lalign/.20685.2.seq Reverse Complement of SEQ ID NO:389 (USPN 5604116
 using matrix file: DNA, gap penalties: -14/-4

  59.8% identity in 97 nt overlap; score:   84 E(10,000):  1e+02

          550        560        570        580        590        600
SEQ     GACTCCGAGTACGTTGTCATGTCCATGCGCATCATGACCCGCATGGGT-ATTGAAGC-GC
        ::: :::  : ::: :  :::::::  : ::   ::: :  :  :: ::::: :  :
Revers GACGACGA--AGGTTAT--TGTCCATCAGGATA--GACATGTCTTCGTCATTGAGGTTGT
     210          220          230          240        250        260

            610        620        630
SEQ     TGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTG
        :: :    : : ::: : :::  ::: : :::::::
Revers TGAAGTCCAGCAGCGGCAGCGGTGGCTGCTTCAGGTG
           270        280        290        300

---

  54.2% identity in 153 nt overlap; score:   75 E(10,000): 5.5e+02

     1490       1500       1510        1520       1530       1540
SEQ     CCCAATGGCA-ATGCTCCCATTCATTGGCT--ACAACGCTGGTGAATACCTGCAGAACTG
        : :::: ::  ::::   : : ::  : ::  ::: : : : ::::: :::  ::   ::
Revers CTCAATTGCTGATGCATTC-TGCAGAGACTTGACAGCACGGTTGAATGCCTCGAGGTTTG
              160        170        180        190        200

        1550       1560       1570       1580       1590       1600
SEQ     GATTGAC-ATGGGTAACAAGGGTGGCGACA-AGATGCCATCCATCTTCCTGGTCAACTGG
        ::  ::: : :: ::       :   :: : : ::: : ::    ::   : :      :
Revers GAC-GACGAAGGTTATTGTCCATCAGGATAGACATGTCTTCGTCATTGAGGTTGTTGAAG
     210        220        230        240        250        260

          1610       1620       1630
SEQ     TTCCGCCGTGGCGAAGATGGACGCTTCCTGTGG
        : : :: : :::   : :::  :::::  ::::
Revers TCCAGCAGCGGCAGCGGTGGCTGCTTCAGGTGG
      270        280        290        300

---

  61.3% identity in 62 nt overlap; score:   74 E(10,000): 6.7e+02

        1650       1660       1670       1680       1690       1700
```

```
SEQ     CGGCGACAACTCTCGCGTTCTGAAGTGGGTCATCGACCGCATCGAAGGCCACGTTGGCGC
        : ::: :: :  : ::    : : :::::: :: ::   ::::::    :: ::: : ::
Revers CAGCGGCAGCGGTGGCTGCTTCAGGTGGGTGAT-GATTTCATCGAT---CATGTTAGAGC
           280       290       300        310          320

SEQ     AG
        ::
Revers AG
     330
```

Back to ISREC bioinformatics group home page

APPENDIX E

lalign output for SEQ ID NO:179 vs. Reverse Complement of SEQ ID NO:389 (USPN 5604116)

[ISREC-Server] Date: Mon Jun 19 19:32:52 Europe/Zurich 2006

---

resetting matrix to DNA ./wwwtmp/lalign/.20685.1.seq : 1953 nt

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
SEQ ID NO:179                                        1953 nt vs.
Reverse Complement of SEQ ID NO:389 (USPN 5604116   339 nt
scoring matrix: DNA, gap penalties: -14/-4
15.1% identity;          Global alignment score: -5757

               10         20         30         40         50         60
./wwwt ATGTGTCCGTTGTCTCACCTAAAGTTTTAACTAGTTCTGTATCTGAAAGCTACGCTAGGG
        :::                                          :::
Revers CTGT----------------------------------------TGA--------------


               70         80         90        100        110        120
./wwwt GGCGAGAACTCTGTCGAATGACACAAAATCTGGAGAAGTAATGACTACTGCTGCAATCAG

Revers ------------------------------------------------------------


              130        140        150        160        170        180
./wwwt GGGCCTTCAGGGCGAGGCGCCGACCAAGAATAAGGAACTGCTGAACTGGATCGCAGACGC
         ::::           ::::
Revers --GCCT----------GCGC----------------------------------------
          10


              190        200        210        220        230        240
./wwwt CGTCGAGCTCTTCCAGCCTGAGGCTGTTGTGTTCGTTGATGGATCCCAGGCTGAGTGGGA
           : ::: : ::             :: :::   ::: ::    ::: : :::
Revers ------GTTCTCCAAG-----------GTTTTC--AGATAGA----AGG-TCAGT----
               20                    30              40


              250        260        270        280        290        300
./wwwt TCGCATGGCGGAGGATCTTGTTGAAGCCGGTACCCTCATCAAGCTCAACGAGGAAAAGCG
                       ::                                 :::: : ::
Revers ----------------TT---------------------------ACGACGGAA----
                                                     50


              310        320        330        340        350        360
./wwwt TCCGAACAGCTACCTAGCTCGTTCCAACCCATCTGACGTTGCGCGCGTTGAGTCCCGCAC

Revers ------------------------------------------------------------


              370        380        390        400        410        420
./wwwt CTTCATCTGCTCCGAGAAGGAAGAAGATGCTGGCCCAACCAACAACTGGGCTCCACCACA
        :::::                                                :::::
```

```
Revers -TTCAT-----------------------------------------------TCCA-----
         60

               430       440       450       460       470       480
./wwwt GGCAATGAAGGACGAAATGTCCAAGCATTACGCTGGTTCCATGAAGGGGCGCACCATGTA
                          :::                                  ::: :
Revers ------------------GTC--------------------------------ACCGT---
                                                              70

               490       500       510       520       530       540
./wwwt CGTCGTGCCTTTCTGCATGGGTCCAATCAGCGATCCGGACCCTAAGCTTGGTGTGCAGCT
               ::::                    :::  :::         ::::
Revers -------CCTT--------------------GATATGGA--------TTGG---------
                                       80                 90

               550       560       570       580       590       600
./wwwt CACTGACTCCGAGTACGTTGTCATGTCCATGCGCATCATGACCCGCATGGGTATTGAAGC
                             :::::                  :: :::::       ::
Revers ----------------------ATGTC-----------------GCGTGGGT------GC
                                                    100

               610       620       630       640       650       660
./wwwt GCTGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTGCCTCCACTCCGTTGGTGCTCC
       :                                            :::             :::
Revers G-------------------------------------GCC------------GTG----
                                                              110

               670       680       690       700       710       720
./wwwt TTTGGAGCCAGGCCAGGAAGACGTTGCATGGCCTTGCAACGACACCAAGTACATCACCCA
             ::::::    :: ::::     :::::
Revers ------GCCAGG---GGCAGAC-----ATGGC----------------------------
                     120             130

               730       740       750       760       770       780
./wwwt GTTCCCAGAGACCAAGGAAATTTGGTCCTACGGTTCCGGCTACGGCGGAAACGCAATCCT
                     :::: ::::           ::
Revers --------------AGGAGATTT----------TT-------------------------
                                       140

               790       800       810       820       830       840
./wwwt GGCAAAGAAGTGCTACGCACTGCGTATCGCATCTGTCATGGCTCGCGAAGAAGGATGGAT
           ::::: ::::                                :::              ::
Revers ----AAGAA-TGCT--------------------------CTCA-------------AT
                  150

               850       860       870       880       890       900
./wwwt GGCTGAGCACATGCTCATCCTGAAGCTGATCAACCCAGAGGGCAAGGCGTACCACATCGC
        :::::     :::  :: ::: ::
Revers TGCTGA-----TGC--ATTCTGCAG-----------------------------------
        160               170

               910       920       930       940       950       960
./wwwt AGCAGCATTCCCATCTGCTTGTGGCAAGACCAACCTCGCCATGATCACTCCAACCATCCC

Revers ------------------------------------------------------------


               970       980       990      1000      1010      1020
./wwwt AGGCTGGACCGCTCAGGTTGTTGGCGACGACATCGCTTGGCTGAAGCTGCGCGAGGACGG
       :: :: ::: :: : :::::                                    :   :
```

```
Revers AGACTTGACAGCAC-GGTTG-----------------------------------AATG
          180        190

              1030      1040      1050      1060      1070      1080
./wwwt CCTCTACGCAGTTAACCCAGAAAATGGTTTCTTCGGTGTTGCTCCAGGCACCAACTACGC
       :::: :                   :::::
Revers CCTCGA------------------GGTTT-----------------------------
       200

              1090      1100      1110      1120      1130      1140
./wwwt ATCCAACCCAATCGCGATGAAGACCATGGAACCAGGCAACACCCTGTTCACCAACGTGGC

Revers ------------------------------------------------------------


              1150      1160      1170      1180      1190      1200
./wwwt ACTCACCGACGACGGCGACATCTGGTGGGAAGGCATGGACGGCGACGCCCCAGCTCACCT
                                           :::::
Revers -------------------------------------GGACG------------------
                                          210

              1210      1220      1230      1240      1250      1260
./wwwt CATTGACTGGATGGGCAACGACTGGACCCCAGAGTCCGACGAAAACGCTGCTCACCCTAA
                                             :::::
Revers ---------------------------------------ACGAA-----------------


              1270      1280      1290      1300      1310      1320
./wwwt CTCCCGTTACTGCGTAGCAATCGACCAGTCCCCAGCAGCAGCACCTGAGTTCAACGACTG

Revers ------------------------------------------------------------


              1330      1340      1350      1360      1370      1380
./wwwt GGAAGGCGTCAAGATCGACGCAATCCTCTTCGGTGGACGTCGCGCAGACACCGTCCCACT

Revers ------------------------------------------------------------


              1390      1400      1410      1420      1430      1440
./wwwt GGTTACCCAGACCTACGACTGGGAGCACGGCACCATGGTTGGTGCACTGCTCGCATCCGG
       :::::                                  : :: ::
Revers GGTTA-----------------------------------TTGTCCA-------------
      220                                        230

              1450      1460      1470      1480      1490      1500
./wwwt TCAGACCGCAGCTTCCGCAGAAGCAAAGGTCGGCACACTCCGCCACGACCCAATGGCAAT
       ::::
Revers TCAG--------------------------------------------------------


              1510      1520      1530      1540      1550      1560
./wwwt GCTCCCATTCATTGGCTACAACGCTGGTGAATACCTGCAGAACTGGATTGACATGGGTAA
                                                   ::: ::::::
Revers ---------------------------------------------GATAGACATG-----
                                                     240

              1570      1580      1590      1600      1610      1620
./wwwt CAAGGGTGGCGACAAGATGCCATCCATCTTCCTGGTCAACTGGTTCCGCCGTGGCGAAGA
                                 :::::    ::::
```

```
Revers ------------------------TCTTC---GTCAT--------------------
                                    250

             1630      1640      1650      1660      1670      1680
./wwwt TGGACGCTTCCTGTGGCCTGGCTTCGGCGACAACTCTCGCGTTCTGAAGTGGGTCATCGA
                                                    ::: :: : :
Revers ---------------------------------------------TGAGGTTGTT------
                                                      260

             1690      1700      1710      1720      1730      1740
./wwwt CCGCATCGAAGGCCACGTTGGCGCAGACGAGACCGTTGTTGGACACACCGCTAAGGCCGA
              :::: :::       ::::
Revers -------GAAGTCCA-------GCAG-----------------------------------
                  270

             1750      1760      1770      1780      1790      1800
./wwwt AGACCTCGACCTCGACGGCCTCGACACCCCAATTGAGGATGTCAAGGAAGCACTGACCGC
                       ::::                                ::: :: ::
Revers ---------------CGGC----------------------------AGCGGTG---GC
                       280

             1810      1820      1830      1840      1850      1860
./wwwt TCCTGCAGAGCAGTGGGCAAACGACGTTGAAGACAACGCCGAGTACCTCACTTTCCTCGG
       : :: :::    :::::          ::: ::                  :::: :::
Revers TGCTTCAG----GTGGG----------TGATGA-----------------TTTCATCG-
     290           300                                    310

             1870      1880      1890      1900      1910      1920
./wwwt ACCACGTGTTCCTGCAGAGGTTCACAGCCAGTTCGATGCTCTGAAGGCCCGCATTTCAGC
       : ::  ::::     ::::          ::::: :                 :::
Revers ATCA--TGTT-----AGAG---------CAGTTAGC---------------CAT------
        320                        330

             1930      1940      1950
./wwwt AGCTCACGCTTAAAGTTCACGCTTAAGAACTGC

Revers --------------------------------
```

Back to ISREC bioinformatics group home page

APPENDIX F

lalign output for SEQ ID NO:179 vs. alleged matched sequence

[ISREC-Server] Date: Mon Jun 19 19:46:12 Europe/Zurich 2006

LALIGN finds the best local alignments between two sequences version 2.0u66 September 1998 Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381 resetting to DNA matrix

```
 Comparison of:
(A) ./wwwtmp/lalign/.11175.1.seq SEQ ID NO:179
(B) ./wwwtmp/lalign/.11175.2.seq alleged matched sequence
 using matrix file: DNA, gap penalties: -14/-4

  84.6% identity in 13 nt overlap; score:   47 E(10,000):  4e+02

           260
SEQ     GATCTTGTTGAAG
        ::  :::::::::
allege GAGGTTGTTGAAG
          10
```

```
  90.9% identity in 11 nt overlap; score:   46 E(10,000): 4.8e+02

            980
SEQ     TCAGGTTGTTG
        : :::::::::
allege TGAGGTTGTTG
           10
```

```
  90.9% identity in 11 nt overlap; score:   46 E(10,000): 4.8e+02

      200
SEQ     TGAGGCTGTTG
        ::::: :::::
allege TGAGGTTGTTG
           10
```

Back to ISREC bioinformatics group home page

APPENDIX G

lalign output for SEQ ID NO:179 vs. alleged matched sequence

[ISREC-Server] Date: Mon Jun 19 19:45:59 Europe/Zurich 2006

---

resetting matrix to DNA ./wwwtmp/lalign/.11175.1.seq : 1953 nt

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
SEQ ID NO:179                                        1953 nt vs.
alleged matched sequence                               18 nt
scoring matrix: DNA, gap penalties: -14/-4
 0.9% identity;          Global alignment score: -7690

               10         20         30         40         50         60
./wwwt ATGTGTCCGTTGTCTCACCTAAAGTTTTAACTAGTTCTGTATCTGAAAGCTACGCTAGGG

allege ------------------------------------------------------------


               70         80         90        100        110        120
./wwwt GGCGAGAACTCTGTCGAATGACACAAAATCTGGAGAAGTAATGACTACTGCTGCAATCAG

allege ------------------------------------------------------------


              130        140        150        160        170        180
./wwwt GGGCCTTCAGGGCGAGGCGCCGACCAAGAATAAGGAACTGCTGAACTGGATCGCAGACGC

allege ------------------------------------------------------------


              190        200        210        220        230        240
./wwwt CGTCGAGCTCTTCCAGCCTGAGGCTGTTGTGTTCGTTGATGGATCCCAGGCTGAGTGGGA

allege ------------------------------------------------------------


              250        260        270        280        290        300
./wwwt TCGCATGGCGGAGGATCTTGTTGAAGCCGGTACCCTCATCAAGCTCAACGAGGAAAAGCG

allege ------------------------------------------------------------


              310        320        330        340        350        360
./wwwt TCCGAACAGCTACCTAGCTCGTTCCAACCCATCTGACGTTGCGCGCGTTGAGTCCCGCAC

allege ------------------------------------------------------------


              370        380        390        400        410        420
./wwwt CTTCATCTGCTCCGAGAAGGAAGAAGATGCTGGCCCAACCAACAACTGGGCTCCACCACA

allege ------------------------------------------------------------
```

```
               430       440       450       460       470       480
./wwwt GGCAATGAAGGACGAAATGTCCAAGCATTACGCTGGTTCCATGAAGGGGCGCACCATGTA

allege ------------------------------------------------------------

               490       500       510       520       530       540
./wwwt CGTCGTGCCTTTCTGCATGGGTCCAATCAGCGATCCGGACCCTAAGCTTGGTGTGCAGCT

allege ------------------------------------------------------------

               550       560       570       580       590       600
./wwwt CACTGACTCCGAGTACGTTGTCATGTCCATGCGCATCATGACCCGCATGGGTATTGAAGC

allege ------------------------------------------------------------

               610       620       630       640       650       660
./wwwt GCTGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTGCCTCCACTCCGTTGGTGCTCC

allege ------------------------------------------------------------

               670       680       690       700       710       720
./wwwt TTTGGAGCCAGGCCAGGAAGACGTTGCATGGCCTTGCAACGACACCAAGTACATCACCCA

allege ------------------------------------------------------------

               730       740       750       760       770       780
./wwwt GTTCCCAGAGACCAAGGAAATTTGGTCCTACGGTTCCGGCTACGGCGGAAACGCAATCCT

allege ------------------------------------------------------------

               790       800       810       820       830       840
./wwwt GGCAAAGAAGTGCTACGCACTGCGTATCGCATCTGTCATGGCTCGCGAAGAAGGATGGAT

allege ------------------------------------------------------------

               850       860       870       880       890       900
./wwwt GGCTGAGCACATGCTCATCCTGAAGCTGATCAACCCAGAGGGCAAGGCGTACCACATCGC

allege ------------------------------------------------------------

               910       920       930       940       950       960
./wwwt AGCAGCATTCCCATCTGCTTGTGGCAAGACCAACCTCGCCATGATCACTCCAACCATCCC

allege ------------------------------------------------------------

               970       980       990      1000      1010      1020
./wwwt AGGCTGGACCGCTCAGGTTGTTGGCGACGACATCGCTTGGCTGAAGCTGCGCGAGGACGG

allege ------------------------------------------------------------
```

```
            1030      1040      1050      1060      1070      1080
./wwwt CCTCTACGCAGTTAACCCAGAAAATGGTTTCTTCGGTGTTGCTCCAGGCACCAACTACGC

allege ------------------------------------------------------------


            1090      1100      1110      1120      1130      1140
./wwwt ATCCAACCCAATCGCGATGAAGACCATGGAACCAGGCAACACCCTGTTCACCAACGTGGC

allege ------------------------------------------------------------


            1150      1160      1170      1180      1190      1200
./wwwt ACTCACCGACGACGGCGACATCTGGTGGGAAGGCATGGACGGCGACGCCCCAGCTCACCT
                                                                  :
allege ------------------------------------------------------------T


            1210      1220      1230      1240      1250      1260
./wwwt CATTGACTGGATGGGCAACGACTGGACCCCAGAGTCCGACGAAAACGCTGCTCACCCTAA
       ::::::
allege CATTGA------------------------------------------------------


            1270      1280      1290      1300      1310      1320
./wwwt CTCCCGTTACTGCGTAGCAATCGACCAGTCCCCAGCAGCAGCACCTGAGTTCAACGACTG

allege ------------------------------------------------------------


            1330      1340      1350      1360      1370      1380
./wwwt GGAAGGCGTCAAGATCGACGCAATCCTCTTCGGTGGACGTCGCGCAGACACCGTCCCACT

allege ------------------------------------------------------------


            1390      1400      1410      1420      1430      1440
./wwwt GGTTACCCAGACCTACGACTGGGAGCACGGCACCATGGTTGGTGCACTGCTCGCATCCGG
                                            :::::
allege --------------------------------------GGTTG------------------
                                             10


            1450      1460      1470      1480      1490      1500
./wwwt TCAGACCGCAGCTTCCGCAGAAGCAAAGGTCGGCACACTCCGCCACGACCCAATGGCAAT

allege ------------------------------------------------------------


            1510      1520      1530      1540      1550      1560
./wwwt GCTCCCATTCATTGGCTACAACGCTGGTGAATACCTGCAGAACTGGATTGACATGGGTAA

allege ------------------------------------------------------------


            1570      1580      1590      1600      1610      1620
./wwwt CAAGGGTGGCGACAAGATGCCATCCATCTTCCTGGTCAACTGGTTCCGCCGTGGCGAAGA

allege ------------------------------------------------------------
```

```
            1630      1640      1650      1660      1670      1680
./wwwt TGGACGCTTCCTGTGGCCTGGCTTCGGCGACAACTCTCGCGTTCTGAAGTGGGTCATCGA

allege ------------------------------------------------------------


            1690      1700      1710      1720      1730      1740
./wwwt CCGCATCGAAGGCCACGTTGGCGCAGACGAGACCGTTGTTGGACACACCGCTAAGGCCGA

allege ------------------------------------------------------------


            1750      1760      1770      1780      1790      1800
./wwwt AGACCTCGACCTCGACGGCCTCGACACCCCAATTGAGGATGTCAAGGAAGCACTGACCGC

allege ------------------------------------------------------------


            1810      1820      1830      1840      1850      1860
./wwwt TCCTGCAGAGCAGTGGGCAAACGACGTTGAAGACAACGCCGAGTACCTCACTTTCCTCGG
                                  ::::::
allege --------------------------TTGAAG----------------------------


            1870      1880      1890      1900      1910      1920
./wwwt ACCACGTGTTCCTGCAGAGGTTCACAGCCAGTTCGATGCTCTGAAGGCCCGCATTTCAGC

allege ------------------------------------------------------------


            1930      1940      1950
./wwwt AGCTCACGCTTAAAGTTCACGCTTAAGAACTGC

allege ---------------------------------
```

Back to ISREC bioinformatics group home page

APPENDIX H

lalign output for Reverse Complement of SEQ ID NO:179 vs. alleged matched sequence

[ISREC-Server] Date: Mon Jun 19 19:45:23 Europe/Zurich 2006

LALIGN finds the best local alignments between two sequences version 2.0u66 September 1998 Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381 resetting to DNA matrix

```
 Comparison of:
(A) ./wwwtmp/lalign/.13439.1.seq Reverse Complement of SEQ ID NO:179
(B) ./wwwtmp/lalign/.13439.2.seq alleged matched sequence
 using matrix file: DNA, gap penalties: -14/-4

  77.8% identity in 18 nt overlap; score:   54 E(10,000): 1.1e+02

           1670      1680
Revers TCGTTGAGCTTGATGAGG
       :: ::::: ::: ::: :
allege TCATTGAGGTTGTTGAAG
               10
```

```
  90.0% identity in 10 nt overlap; score:   41 E(10,000): 1.2e+03

       750
Revers TCAATGAGGT
       ::: ::::::
allege TCATTGAGGT
               10
```

```
  81.8% identity in 11 nt overlap; score:   37 E(10,000): 2.4e+03

        820
Revers ACGTTGGTGAA
       : :::: ::::
allege AGGTTGTTGAA
         10
```

Back to ISREC bioinformatics group home page

APPENDIX I

lalign output for Reverse Complement of SEQ ID NO:179 vs. alleged matched sequence

[ISREC-Server] Date: Mon Jun 19 19:45:36 Europe/Zurich 2006

---

resetting matrix to DNA ./wwwtmp/lalign/.13439.1.seq : 1953 nt

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
Reverse Complement of SEQ ID NO:179                1953 nt vs.
alleged matched sequence                             18 nt
scoring matrix: DNA, gap penalties: -14/-4
 0.9% identity;          Global alignment score: -7699

               10         20         30         40         50         60
./wwwt GCAGTTCTTAAGCGTGAACTTTAAGCGTGAGCTGCTGAAATGCGGGCCTTCAGAGCATCG

allege ------------------------------------------------------------


               70         80         90        100        110        120
./wwwt AACTGGCTGTGAACCTCTGCAGGAACACGTGGTCCGAGGAAAGTGAGGTACTCGGCGTTG

allege ------------------------------------------------------------


              130        140        150        160        170        180
./wwwt TCTTCAACGTCGTTTGCCCACTGCTCTGCAGGAGCGGTCAGTGCTTCCTTGACATCCTCA

allege ------------------------------------------------------------


              190        200        210        220        230        240
./wwwt ATTGGGGTGTCGAGGCCGTCGAGGTCGAGGTCTTCGGCCTTAGCGGTGTGTCCAACAACG

allege ------------------------------------------------------------


              250        260        270        280        290        300
./wwwt GTCTCGTCTGCGCCAACGTGGCCTTCGATGCGGTCGATGACCCACTTCAGAACGCGAGAG

allege ------------------------------------------------------------


              310        320        330        340        350        360
./wwwt TTGTCGCCGAAGCCAGGCCACAGGAAGCGTCCATCTTCGCCACGGCGGAACCAGTTGACC

allege ------------------------------------------------------------


              370        380        390        400        410        420
./wwwt AGGAAGATGGATGGCATCTTGTCGCCACCCTTGTTACCCATGTCAATCCAGTTCTGCAGG
```

```
allege ------------------------------------------------------------

               430       440       450       460       470       480
./wwwt TATTCACCAGCGTTGTAGCCAATGAATGGGAGCATTGCCATTGGGTCGTGGCGGAGTGTG

allege ------------------------------------------------------------

               490       500       510       520       530       540
./wwwt CCGACCTTTGCTTCTGCGGAAGCTGCGGTCTGACCGGATGCGAGCAGTGCACCAACCATG

allege ------------------------------------------------------------

               550       560       570       580       590       600
./wwwt GTGCCGTGCTCCCAGTCGTAGGTCTGGGTAACCAGTGGGACGGTGTCTGCGCGACGTCCA

allege ------------------------------------------------------------

               610       620       630       640       650       660
./wwwt CCGAAGAGGATTGCGTCGATCTTGACGCCTTCCCAGTCGTTGAACTCAGGTGCTGCTGCT

allege ------------------------------------------------------------

               670       680       690       700       710       720
./wwwt GGGGACTGGTCGATTGCTACGCAGTAACGGGAGTTAGGGTGAGCAGCGTTTTCGTCGGAC

allege ------------------------------------------------------------

               730       740       750       760       770       780
./wwwt TCTGGGGTCCAGTCGTTGCCCATCCAGTCAATGAGGTGAGCTGGGGCGTCGCCGTCCATG
                                    ::: ::::::
allege ----------------------------TCATTGAGGT----------------------
                                          10

               790       800       810       820       830       840
./wwwt CCTTCCCACCAGATGTCGCCGTCGTCGGTGAGTGCCACGTTGGTGAACAGGGTGTTGCCT
                                                                 :::::
allege ------------------------------------------------------TGTTG---

               850       860       870       880       890       900
./wwwt GGTTCCATGGTCTTCATCGCGATTGGGTTGGATGCGTAGTTGGTGCCTGGAGCAACACCG

allege ------------------------------------------------------------

               910       920       930       940       950       960
./wwwt AAGAAACCATTTTCTGGGTTAACTGCGTAGAGGCCGTCCTCGCGCAGCTTCAGCCAAGCG
       :::
allege AAG---------------------------------------------------------

               970       980       990      1000      1010      1020
./wwwt ATGTCGTCGCCAACAACCTGAGCGGTCCAGCCTGGGATGGTTGGAGTGATCATGGCGAGG
```

```
allege ------------------------------------------------------------

            1030       1040       1050       1060       1070       1080
./wwwt TTGGTCTTGCCACAAGCAGATGGGAATGCTGCTGCGATGTGGTACGCCTTGCCCTCTGGG

allege ------------------------------------------------------------

            1090       1100       1110       1120       1130       1140
./wwwt TTGATCAGCTTCAGGATGAGCATGTGCTCAGCCATCCATCCTTCTTCGCGAGCCATGACA

allege ------------------------------------------------------------

            1150       1160       1170       1180       1190       1200
./wwwt GATGCGATACGCAGTGCGTAGCACTTCTTTGCCAGGATTGCGTTTCCGCCGTAGCCGGAA

allege ------------------------------------------------------------

            1210       1220       1230       1240       1250       1260
./wwwt CCGTAGGACCAAATTTCCTTGGTCTCTGGGAACTGGGTGATGTACTTGGTGTCGTTGCAA

allege ------------------------------------------------------------

            1270       1280       1290       1300       1310       1320
./wwwt GGCCATGCAACGTCTTCCTGGCCTGGCTCCAAAGGAGCACCAACGGAGTGGAGGCACCTG

allege ------------------------------------------------------------

            1330       1340       1350       1360       1370       1380
./wwwt ACGAAGCTGCCGTTCGCGCCGATCTTGTCCAGCGCTTCAATACCCATGCGGGTCATGATG

allege ------------------------------------------------------------

            1390       1400       1410       1420       1430       1440
./wwwt CGCATGGACATGACAACGTACTCGGAGTCAGTGAGCTGCACACCAAGCTTAGGGTCCGGA

allege ------------------------------------------------------------

            1450       1460       1470       1480       1490       1500
./wwwt TCGCTGATTGGACCCATGCAGAAAGGCACGACGTACATGGTGCGCCCCTTCATGGAACCA

allege ------------------------------------------------------------

            1510       1520       1530       1540       1550       1560
./wwwt GCGTAATGCTTGGACATTTCGTCCTTCATTGCCTGTGGTGGAGCCCAGTTGTTGGTTGGG

allege ------------------------------------------------------------

            1570       1580       1590       1600       1610       1620
./wwwt CCAGCATCTTCTTCCTTCTCGGAGCAGATGAAGGTGCGGGACTCAACGCGCGCAACGTCA
```

```
allege ------------------------------------------------------------

             1630       1640       1650       1660       1670       1680
./wwwt GATGGGTTGGAACGAGCTAGGTAGCTGTTCGGACGCTTTTCCTCGTTGAGCTTGATGAGG

allege ------------------------------------------------------------

             1690       1700       1710       1720       1730       1740
./wwwt GTACCGGCTTCAACAAGATCCTCCGCCATGCGATCCCACTCAGCCTGGGATCCATCAACG

allege ------------------------------------------------------------

             1750       1760       1770       1780       1790       1800
./wwwt AACACAACAGCCTCAGGCTGGAAGAGCTCGACGGCGTCTGCGATCCAGTTCAGCAGTTCC

allege ------------------------------------------------------------

             1810       1820       1830       1840       1850       1860
./wwwt TTATTCTTGGTCGGCGCCTCGCCCTGAAGGCCCCTGATTGCAGCAGTAGTCATTACTTCT

allege ------------------------------------------------------------

             1870       1880       1890       1900       1910       1920
./wwwt CCAGATTTTGTGTCATTCGACAGAGTTCTCGCCCCCTAGCGTAGCTTTCAGATACAGAAC

allege ------------------------------------------------------------

             1930       1940       1950
./wwwt TAGTTAAAACTTTAGGTGAGACAACGGACACAT

allege ---------------------------------
```

Back to ISREC bioinformatics group home page

Appendix J

lalign output for SEQ ID NO:179 vs. USPN 6294328 nts. 251891-253705

[ISREC-Server] Date: Tue Jun 20 19:08:21 Europe/Zurich 2006

LALIGN finds the best local alignments between two sequences version 2.0u66 September 1998 Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381 resetting to DNA matrix

```
 Comparison of:
(A) ./wwwtmp/lalign/.15652.1.seq SEQ ID NO:179
(B) ./wwwtmp/lalign/.15652.2.seq USPN 6294328 nts. 251891-253705
 using matrix file: DNA, gap penalties: -14/-4

  68.6% identity in 1833 nt overlap; score: 3781 E(10,000): 3.2e-306

              110        120        130        140        150        160
SEQ     ATGACTACTGCTGCAATCAGGGGCCTTCAGGGCGAGGCGCCGACCAAGAATAAGGAACTG
        :::::  : ::  : :::   ::   :: ::      :::::::: ::  :  :::   ::
USPN    ATGACCTCAGCGACCATCCCCGG---TCTGGATACCGCGCCGACGAATCACCAGGGGTTG
                10         20            30         40         50

              170        180        190        200        210        220
SEQ     CTGAACTGGATCGCAGACGCCGTCGAGCTCTTCCAGCCTGAGGCTGTTGTGTTCGTTGAT
        :::  :::: ::: ::: : :: :::::::  :::::: ::    :: :: ::: :::
USPN    CTGTCCTGGGTCGAAGAGGTCGCCGAGCTCACCCAGCCGGACCGGGTGGTCTTCACTGAC
         60         70         80         90        100        110

              230        240        250        260        270        280
SEQ     GGATCCCAGGCTGAGTGGGATCGCATGGCGGAGGATCTTGTTGAAGCCGGTACCCTCATC
        :: ::  : :  ::::   : ::  :    ::  : :: :: :: ::::: ::  :::::
USPN    GGCTCGGAAGAAGAGTTCCAGCGGCTCTGCGATCAGCTAGTCGAGGCCGGCACGTTCATC
        120        130        140        150        160        170

              290        300        310        320        330        340
SEQ     AAGCTCAACGAGGAAAAGCGTCCGAACAGCTACCTAGCTCGTTCCAACCCATCTGACGTT
        : :::::::   :: ::::    ::::  :::::: ::    ::  : :: :: :: ::
USPN    AGGCTCAACCCCGAGAAGCACAAGAACTCCTACCTGGCATTGTCGGATCCGTCCGATGTC
        180        190        200        210        220        230

              350        360        370        380        390        400
SEQ     GCGCGCGTTGAGTCCCGCACCTTCATCTGCTCCGAGAAGGAAGAAGATGCTGGCCCAACC
        ::::: :: ::::: :: :: : ::::::::: : ::: ::    :: :: ::::: :::
USPN    GCGCGGGTGGAGTCGCGGACGTACATCTGCTCGGCGAAAGAGATCGACGCCGGCCCCACC
        240        250        260        270        280        290

              410        420        430        440        450
SEQ     AACAACTGGGCTCCACCACAG-GCAATGAAGGACGAAATGTCCAAGCATTACGCTGGTTC
        :::::::::  :  : : : : : :::  :  :   :::    : :  :::   ::::
USPN    AACAACTGGA-TGGATCCCGGCGAAATGCGGTCCATCATGAAAGACCTGTACCGGGGTTG
        300        310        320        330        340        350

     460        470        480        490        500        510
SEQ     CATGAAGGGGCGCACCATGTACGTCGTGCCTTTCTGCATGGGTCCAATCAGCGATCCGGA
        ::::   :::::::::::::: :: ::::: ::::: ::::: ::  :  :::    :::
```

```
USPN   CATGCGCGGGCGCACCATGTATGTGGTGCCGTTCTGTATGGGACCGCTGGGCGCCGAGGA
        360        370        380        390        400        410

     520        530        540        550        560        570
SEQ    CCCTAAGCTTGGTGTGCAGCTCACTGACTCCGAGTACGTTGTCATGTCCATGCGCATCAT
       ::: :: ::::::::: :: :::: :::::::::::::: ::: : :::::::::: :::
USPN   CCCCAAACTTGGTGTGGAGATCACCGACTCCGAGTACGTCGTCGTCTCCATGCGCACCAT
        420        430        440        450        460        470

     580        590        600        610        620        630
SEQ    GACCCGCATGGGTATTGAAGCGCTGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTG
       :::::: ::::: :  :  :::::::: :: :: ::::    ::::    ::: :::: :
USPN   GACCCGGATGGGCAAGGCCGCGCTGGAGAAAATGGGCGACGACGGTTTCTTTGTCAAGGC
        480        490        500        510        520        530

     640        650        660        670        680        690
SEQ    CCTCCACTCCGTTGGTGCTCCTTTGGAGCCAGGCCAGGAAGACGTTGCATGGCCTTGCAA
        :: ::::: :: :: :: ::  :::: :: :::::  : ::::: :: ::::: ::::
USPN   GCTGCACTCGGTCGGCGCGCCGCTGGAACCGGGCCAAAAGGACGTGGCCTGGCCCTGCAG
        540        550        560        570        580        590

     700        710        720        730        740        750
SEQ    CGACACCAAGTACATCACCCAGTTCCCAGAGACCAAGGAAATTTGGTCCTACGGTTCCGG
       ::: ::::::::::::::::: ::::: ::::::  ::: :: :::  :::::: :: ::
USPN   CGAAACCAAGTACATCACCCACTTCCCGGAGACCCGGGAGATCTGGAGCTACGGCTCGGG
        600        610        620        630        640        650

     760        770        780        790        800        810
SEQ    CTACGGCGGAAACGCAATCCTGGCAAAGAAGTGCTACGCACTGCGTATCGCATCTGTCAT
       ::::::::: :::::  : ::::  :::::::::::: :::::::::::: :: :  ::
USPN   CTACGGCGGCAACGCGTTGCTGGGCAAGAAGTGCTACTCACTGCGTATCGCGTCGGCGAT
        660        670        680        690        700        710

     820        830        840        850        860        870
SEQ    GGCTCGCGAAGAAGGATGGATGGCTGAGCACATGCTCATCCTGAAGCTGATCAACCCAGA
       ::: : ::: :: :: ::: :::: :::::::::: ::::: ::::::::    :: ::
USPN   GGCCCACGATGAGGGCTGGCTGGCCGAGCACATGCTGATCCTCAAGCTGATTTCGCCGGA
        720        730        740        750        760        770

     880        890        900        910        920        930
SEQ    GGGCAAGGCGTACCACATCGCAGCAGCATTCCCATCTGCTTGTGGCAAGACCAACCTCGC
       :  :::::: ::: :: :::: :: :::::::: :: :: ::::::::::::::::: ::
USPN   GAACAAGGCTTACTACTTCGCGGCCGCATTCCCGTCGGCGTGTGGCAAGACCAACCTGGC
        780        790        800        810        820        830

     940        950        960        970        980        990
SEQ    CATGATCACTCCAACCATCCCAGGCTGGACCGCTCAGGTTGTTGGCGACGACATCGCTTG
       ::: :    :::::::::: ::::::   ::  ::    : :: :::::::::: ::
USPN   GATGCTGCAGCCAACCATCCCCGGCTGGCGTGCGGAGACACTCGGAGACGACATCGCATG
        840        850        860        870        880        890

    1000       1010       1020       1030       1040       1050
SEQ    GCTGAAGCTGCGCGAGGACGG---CCTCTACGCAGTTAACCCAGAAAATGGTTTCTTCGG
       : ::    :  :: ::::::::   ::: ::::: :: ::::: :::   :: ::::::::
USPN   GATGCGATTTGGCAAGGACGGTCGCCTGTACGCCGTCAACCCGGAATTCGGCTTCTTCGG
        900        910        920        930        940        950

       1060       1070       1080       1090       1100       1110
SEQ    TGTTGCTCCAGGCACCAACTACGCATCCAACCCAATCGCGATGAAGACCATGGAACCAGG
        :: :: :: ::::::::::      :: ::::: : ::: :::   ::::: :   : ::
```

```
USPN   GGTGGCGCCGGGCACCAACTGGAAGTCGAACCCTAACGCCATGCGCACCATTGCCGCCGG
        960        970        980        990       1000       1010

       1120       1130       1140       1150       1160       1170
SEQ    CAACACCCTGTTCACCAACGTGGCACTCACCGACGACGGCGACATCTGGTGGGAAGGCAT
       ::::::  :::::::::: :: ::::::::::::::::::::: : :::::::: ::: :
USPN   CAACACGGTGTTCACCAATGTCGCACTCACCGACGACGGCGACGTGTGGTGGGAGGGCCT
       1020       1030       1040       1050       1060       1070

       1180       1190       1200       1210       1220          1230
SEQ    GGACGGCGACGCCCCAGCTCACCTCATTGACTGGATGGGCAACGACTGG-AC--CCCAGA
       ::: :::::: :: :::  ::::: :: ::::::: ::::::::::::: ::  ::  ::
USPN   GGAAGGCGAC-CCGCAG--CACCTGATCGACTGGAAGGGCAACGACTGGTACTTCCGCGA
       1080        1090         1100       1110       1120       1130

          1240       1250       1260       1270       1280       1290
SEQ    GTCCGACGAAAACGCTGCTCACCCTAACTCCCGTTACTGCGTAGCAATCGACCAGTCCCC
       : : ::     :: :: :: ::::: :::::::: ::::::  : : ::     :::: :::
USPN   GACGGAAACCAATGCGGCACACCCGAACTCCCGGTACTGCACACCGATGTCGCAGTGCCC
          1140       1150       1160       1170       1180       1190

          1300       1310       1320       1330       1340       1350
SEQ    AGCAGCAGCACCTGAGTTCAACGACTGGGAAGGCGTCAAGATCGACGCAATCCTCTTCGG
             :: :: ::::   : :::  : : :::::::  ::::   :  ::::: :::::
USPN   GATCCTGGCCCCCGAGTGGGATGACCCGCAGGGCGTCCCGATCTCGGGGATCCTGTTCGG
          1200       1210       1220       1230       1240       1250

          1360        1370       1380       1390       1400       1410
SEQ    TGGACGTCGCGCAGAC-ACCGTCCCACTGGTTACCCAGACCTACGACTGGGAGCACGGCA
        :: :: :::  :::: :: :: :: ::::: ::: :: :   ::::::: :::::::
USPN   CGGCCGCCGCA-AGACCACGGTTCCGCTGGTCACCGAGGCGCGCGACTGGCAGCACGGGG
          1260       1270       1280       1290       1300       1310

           1420       1430       1440       1450       1460       1470
SEQ    CCATGGTTGGTGCACTGCTCGCATCCGGTCAGACCGCAGCTTCCGCAGAAGCAAAGGTCG
          :  : :::::    :: :    ::  :::::::: ::  :::   : :  :::::::
USPN   TGTTCATCGGTGCGACCCTGGGTAGCGAGCAGACCGCCGCGGCCG---AGGGCAAGGTCG
           1320       1330       1340       1350          1360

           1480       1490       1500       1510       1520       1530
SEQ    GCACACTCCGCCACGACCCAATGGCAATGCTCCCATTCATTGGCTACAACGCTGGTGAAT
       :::   : :::: :::::: ::::: ::::: :: ::  : :::::::::: ::: :: :
USPN   GCAATGTGCGCCGCGACCCGATGGCCATGCTGCCGTTTTTGGGCTACAACGTTGGGGACT
    1370       1380       1390       1400       1410       1420

           1540       1550       1560       1570          1580
SEQ    ACCTGCAGAACTGGATTGACATGGGTAACAAGGGTGGCGA---CAAGATGCCATCCATCT
       :: : ::: :::::::  :: :::: ::  : :  :  ::   :::: ::::     : :
USPN   ACTTCCAGCACTGGATCAACCTGGGCAAGCACGCCGATGAGTCCAAGCTGCCCAAGGTGT
    1430       1440       1450       1460       1470       1480

    1590       1600       1610       1620       1630       1640
SEQ    TCCTGGTCAACTGGTTCCGCCGTGGCGAAGATGGACGCTTCCTGTGGCCTGGCTTCGGCG
       :: : :::::::::::::: :: :: :: :: :: :::::::::::::: ::::::::::
USPN   TCTTCGTCAACTGGTTCCGTCGCGGTGACGACGGTCGCTTCCTGTGGCCGGGCTTCGGCG
    1490       1500       1510       1520       1530       1540

    1650       1660       1670       1680       1690       1700
SEQ    ACAACTCTCGCGTTCTGAAGTGGGTCATCGACCGCATCGAAGGCCACGTTGGCGCAGACG
       : :::   :: :: ::::::::: :: :::: ::::::::   : : :  ::::  : ::
```

```
USPN   AGAACAGCCGGGTGCTGAAGTGGATCGTCGATCGCATCGAGCACAAGGCCGGCGGTG-CG
    1550       1560       1570       1580       1590       1600

    1710       1720       1730          1740       1750       1760
SEQ    AG-ACCGTTGTTGGACACACCGCTAAGGCCG---AAGACCTCGACCTCGACGGCCTCGAC
       :  :::    : ::   ::::: :   ::::   : ::: : ::::: ::::: :: :::
USPN   ACCACCCCGATCGG---CACCGTTCCCGCCGTGGAGGACTTGGACCTGGACGGACTGGAC
     1610       1620          1630       1640       1650       1660

        1770       1780       1790       1800       1810       1820
SEQ    ACCCCAATTGAGGATGTCAAGGAAGCACTGACCGCTCCTGCAGAGCAGTGGGCAAACGAC
         :       :  :::::     :  :: ::: : :    ::: ::  : :::     : ::
USPN   GTCGACGCCGCCGATGTAGCCGCGGCGCTGGCAGTCGATGCCGATGAATGGCGTCAGGAA
        1670       1680       1690       1700       1710       1720

        1830       1840       1850       1860       1870         1880
SEQ    GTTGAAGACAACGCCGAGTACCTCACTTTCCTCGGACCACGTGTTCCTGCAGA--GGT-T
        :        : ::  :: :  ::     ::: ::::    :: :   :::: ::  ::: :
USPN   CTGCCGCTGATCGAAGAATGGCTGCAGTTCGTCGG----CGAGAAGCTGCCGACCGGTGT
        1730       1740       1750       1760          1770       1780

            1890       1900       1910
SEQ    CACAGCC-AGTTCGATGCTCTGAAGGCCCGCAT
       :: ::   ::::::: :: :::::::  ::: :
USPN   CAAAGATGAGTTCGACGCCCTGAAGGAGCGCCT
            1790       1800       1810
```

55.6% identity in 471 nt overlap; score: 134 E(10,000): 0.04

```
            820        830        840        850        860          870
SEQ    CTGTCATGGCTCGCGAAGAAGGATGGATGGCTGAGCACATGCTCATCCTGAA--GCTGAT
       ::::: ::: ::  ::::: :       : :: :::: ::  : :: :: ::    : :: :
USPN   CTGTCCTGGGTC--GAAGAGG------TCGCCGAGCTCA--CCCAGCCGGACCGGGTGGT
        60           70               80           90        100

               880        890        900        910        920        930
SEQ    CAACCCAGAGGGCAAGGCGTACCACATCGCAGCAGCATTCCCATCTGCTTGTGGCAAGAC
       :  : : :: :::  ::   :  :  :: :::: ::  : : ::: ::: :: : : : :
USPN   CTTCACTGACGGCTCGGAAGAAGAGTTC-CAGCGGCTCTGCGATCAGCTAGTCG-AGGCC
       110        120        130        140        150        160

                940        950        960          970        980
SEQ    --CAACCTCGCCATGATCA-CTCCAACCATCCCAGG--CTGG-ACCGCTCAGGTTGTTGG
         ::   ::  :: : ::: : :: :  : : :: :  ::   :::   ::  :::: ::
USPN   GGCACGTTCATCAGGCTCAACCCCGAGAAGCACAAGAACTCCTACCTGGCA--TTGTCGG
         170        180        190        200        210          220

               990       1000       1010       1020       1030
SEQ    ---CGAC-GACATCGCTTGGCTGAAGCTGCGCGAGGACGGCC-TCTACGCAGTTAACCCA
          :: : ::  ::::  :: :: ::  :::    ::::  : ::: : : :  ::   :
USPN   ATCCGTCCGATGTCGCGCGGGTGGAGTCGCG----GACGTACATCTGCTCGGCGAA---A
           230        240        250           260        270

    1040       1050       1060       1070       1080       1090
SEQ    GAAAATGGTTTCTTCGGTGTTGCTCCAGGCACCAACTACGCATCCAACCCAATCGCGATG
       :: :      ::  ::  :  :: :::  :: :::::  : ::  : :::   ::  :::
USPN   GAGA------TCGACGCCG--GCCCCAC-CAACAACT--GGATGGATCCCGG-CGAAATG
        280               290        300          310        320
```

```
    1100       1110       1120       1130       1140       1150
SEQ     AAGACCATGGAACCAGGCAACACCCTGTTCACCAACGTGGCACTCACCGACG-ACGGCGA
          : ::::      :: : :: ::: :::  :::    :: :::  : : : :: ::   :
USPN    CGGTCCAT-----CATGAAAGACC-TGT--ACCGGGGTTGCATGCGCGGGCGCACCATGT
           330              340           350        360        370

     1160        1170        1180       1190       1200       1210
SEQ     CATCTGGTG--GGAAGGCATGG-ACGGCGACGCCCCAGCTCACCTCATTGACTGGATGGG
         :: :::::  :     : :::: :: ::   :: :: :   ::: ::    ::: : :: :
USPN    -ATGTGGTGCCGTTCTGTATGGGACCGCTGGGCGCC-GAGGACCCCA--AACTTGGTGTG
          380        390        400        410         420          430

          1220       1230       1240         1250       1260
SEQ     CA-ACGACTGGACCCCAGAGTCCGACGAAAACGC--TGCTCACCCTAACTC
         : :  :: : :: :: :::: :: ::     : :  ::: :::: : :: :
USPN    GAGATCACCG-ACTCC-GAGTACGTCGTCGTCTCCATGCGCACCATGACCC
             440          450        460        470        480
```

53.4% identity in 266 nt overlap; score: 114 E(10,000): 1.9

```
          1560       1570          1580       1590       1600
SEQ     TGGGTAACAAGGGTGGCGACAAGAT---GCCATCCATCTTCCTGGTCAAC--------TG
        ::::  :::: : ::: ::: :  :    ::  :  :::  ::::: :::          ::
USPN    TGGGCTACAACGTTGGGGACTACTTCCAGCACTGGATCAACCTGGGCAAGCACGCCGATG
     1410       1420       1430       1440       1450       1460

           1610       1620       1630       1640       1650       1660
SEQ     GTTCCGCCGTGGCGAAGATGGACG-CTTCCTGTGGCCTGGCTTCGGCGACAACTCTCGCG
          :::     :: : ::: ::  :  : ::   :::     :   ::: ::: ::  ::::
USPN    AGTCCAAGCTGCCCAAGGTGTTCTTCGTCAACTGGTTCCGTCGCGGTGACGACGGTCGCT
     1470       1480       1490       1500       1510       1520

             1670       1680       1690       1700       1710        1720
SEQ     TTCTGAAGTGGGTCATCGACCGCATCGAAGGCCACGTTGGCGCAGACGAG-ACCGTTGTT
        : :::  :  :: : ::: :          ::  ::     :: :: :: : : : ::: : :
USPN    TCCTGTGGCCGGGCTTCGGCG-------AGAACAGCCGGGTGCTGAAGTGGATCGTCGAT
     1530       1540              1550       1560       1570       1580

             1730        1740         1750        1760       1770
SEQ     GGACACACCGCT-AAGGCCGAAGA--CCTCGACCTCGA-CGGCCTCGACACCCCAATTGA
         : ::     ::  :::::::  :    :  : ::: ::: ::::  ::    :: :  : ::
USPN    CG-CATCGAGCACAAGGCCGGCGGTGCGACCACCCCGATCGGCACCGTTCCCGCCGTGGA
              1590       1600       1610       1620       1630       1640

        1780          1790
SEQ     GGATGTCAA---GGAAGCACTGACCG
        :::  :  :   ::: : ::::  ::
USPN    GGACTTGGACCTGGACGGACTGGACG
              1650       1660
```

Back to ISREC bioinformatics group home page

Appendix K

lalign output for SEQ ID NO:179 vs. USPN 6294328 nts. 251891-253705

[ISREC-Server] Date: Tue Jun 20 18:53:01 Europe/Zurich 2006

---

resetting matrix to DNA ./wwwtmp/lalign/.13200.1.seq : 1953 nt

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989) 4:11-17
SEQ ID NO:179                                        1953 nt vs.
USPN 6294328 nts. 251891-253705                      1815 nt
scoring matrix: DNA, gap penalties: -14/-4
64.2% identity;          Global alignment score: 3236

               10         20         30         40         50         60
./wwwt ATGTGTCCGTTGTCTCACCTAAAGTTTTAACTAGTTCTGTATCTGAAAGCTACGCTAGGG
       :::              :::: :                              :: ::  :
USPN   ATG-------------ACCTCA--------------------------GCGACCATCCCC
                                                       10         20

               70         80         90        100        110        120
./wwwt GGCGAGAACTCTGTCGAATGACACAAAATCTGGAGAAGTAATGACTACTGCTGCAATCAG
       ::                            ::::::           ::: ::
USPN   GG--------------------------TCTGGA-----------TACCGC---------
                                                    30

              130        140        150        160        170        180
./wwwt GGGCCTTCAGGGCGAGGCGCCGACCAAGAATAAGGAACTGCTGAACTGGATCGCAGACGC
                        :::::: ::  :  :::   :::::  :::: ::: ::: :
USPN   -----------------GCCGACGAATCACCAGGGGTTGCTGTCCTGGGTCGAAGAGGT
                           40         50         60         70

              190        200        210        220        230        240
./wwwt CGTCGAGCTCTTCCAGCCTGAGGCTGTTGTGTTCGTTGATGGATCCCAGGCTGAGTGGGA
       :: :::::::  :::::: ::    :: :: :::  ::: :: ::  : :  ::::   :
USPN   CGCCGAGCTCACCCAGCCGGACCGGGTGGTCTTCACTGACGGCTCGGAAGAAGAGTTCCA
        80         90        100        110        120        130

              250        260        270        280        290        300
./wwwt TCGCATGGCGGAGGATCTTGTTGAAGCCGGTACCCTCATCAAGCTCAACGAGGAAAAGCG
        ::  :    ::  : :: :: :: ::::: ::  :::::: :::::::   :: ::::
USPN   GCGGCTCTGCGATCAGCTAGTCGAGGCCGGCACGTTCATCAGGCTCAACCCCGAGAAGCA
       140        150        160        170        180        190

              310        320        330        340        350        360
./wwwt TCCGAACAGCTACCTAGCTCGTTCCAACCCATCTGACGTTGCGCGCGTTGAGTCCCGCAC
          ::::  :::::: ::    ::  : :: :: :: :: ::::: :: ::::: :: ::
USPN   CAAGAACTCCTACCTGGCATTGTCGGATCCGTCCGATGTCGCGCGGGTGGAGTCGCGGAC
       200        210        220        230        240        250

              370        380        390        400        410        420
./wwwt CTTCATCTGCTCCGAGAAGGAAGAAGATGCTGGCCCAACCAACAACTGGGCTCCACCACA
        : ::::::::: : ::: ::    :: :: ::::: ::::::::::::  :  : : :
USPN   GTACATCTGCTCGGCGAAAGAGATCGACGCCGGCCCCACCAACAACTGGA-TGGATCCCG
```

```
       260        270        280        290        300        310

               430        440        450        460        470
./wwwt G-GCAATGAAGGACGAAATGTCCAAGCATTACGCTGGTTCCATGAAGGGGCGCACCATGT
       : : :::: :  :  :::    : :  :::   :::: ::::   ::::::::::::
USPN   GCGAAATGCGGTCCATCATGAAAGACCTGTACCGGGGTTGCATGCGCGGGCGCACCATGT
        320        330        340        350        360        370

     480        490        500        510        520        530
./wwwt ACGTCGTGCCTTTCTGCATGGGTCCAATCAGCGATCCGGACCCTAAGCTTGGTGTGCAGC
       : :: ::::: ::::: ::::: ::  :  :::    :::::: :: ::::::::: ::
USPN   ATGTGGTGCCGTTCTGTATGGGACCGCTGGGCGCCGAGGACCCCAAACTTGGTGTGGAGA
        380        390        400        410        420        430

     540        550        560        570        580        590
./wwwt TCACTGACTCCGAGTACGTTGTCATGTCCATGCGCATCATGACCCGCATGGGTATTGAAG
       :::: :::::::::::::: ::: : :::::::::: ::::::::: ::::: :  :  :
USPN   TCACCGACTCCGAGTACGTCGTCGTCTCCATGCGCACCATGACCCGGATGGGCAAGGCCG
        440        450        460        470        480        490

     600        610        620        630        640        650
./wwwt CGCTGGACAAGATCGGCGCGAACGGCAGCTTCGTCAGGTGCCTCCACTCCGTTGGTGCTC
       ::::::: :: :: ::::   ::::   ::: :::: :   :: ::::: :: :: :: :
USPN   CGCTGGAGAAAATGGGCGACGACGGTTTCTTTGTCAAGGCGCTGCACTCGGTCGGCGCGC
        500        510        520        530        540        550

     660        670        680        690        700        710
./wwwt CTTTGGAGCCAGGCCAGGAAGACGTTGCATGGCCTTGCAACGACACCAAGTACATCACCC
       :  :::: :: :::::  : ::::: :: ::::: :::: ::: ::::::::::::::::
USPN   CGCTGGAACCGGGCCAAAAGGACGTGGCCTGGCCCTGCAGCGAAACCAAGTACATCACCC
        560        570        580        590        600        610

     720        730        740        750        760        770
./wwwt AGTTCCCAGAGACCAAGGAAATTTGGTCCTACGGTTCCGGCTACGGCGGAAACGCAATCC
       : ::::: :::::::  ::: :: :::  :::::: :: :::::::::: :::::  : :
USPN   ACTTCCCGGAGACCCGGGAGATCTGGAGCTACGGCTCGGGCTACGGCGGCAACGCGTTGC
        620        630        640        650        660        670

     780        790        800        810        820        830
./wwwt TGGCAAAGAAGTGCTACGCACTGCGTATCGCATCTGTCATGGCTCGCGAAGAAGGATGGA
       :::  :::::::::::: :::::::::::: :: :  ::::: : ::: :: :: :::
USPN   TGGGCAAGAAGTGCTACTCACTGCGTATCGCGTCGGCGATGGCCCACGATGAGGGCTGGC
        680        690        700        710        720        730

     840        850        860        870        880        890
./wwwt TGGCTGAGCACATGCTCATCCTGAAGCTGATCAACCCAGAGGGCAAGGCGTACCACATCG
       :::: :::::::::: ::::: ::::::::    :: :::  :::::: ::: :: :::
USPN   TGGCCGAGCACATGCTGATCCTCAAGCTGATTTCGCCGGAGAACAAGGCTTACTACTTCG
        740        750        760        770        780        790

     900        910        920        930        940        950
./wwwt CAGCAGCATTCCCATCTGCTTGTGGCAAGACCAACCTCGCCATGATCACTCCAACCATCC
       : :: :::::::: :: :: ::::::::::::::::: :: ::: :    ::::::::::
USPN   CGGCCGCATTCCCGTCGGCGTGTGGCAAGACCAACCTGGCGATGCTGCAGCCAACCATCC
        800        810        820        830        840        850

     960        970        980        990       1000       1010
./wwwt CAGGCTGGACCGCTCAGGTTGTTGGCGACGACATCGCTTGGCTGAAGCTGCGCGAGGACG
       : ::::::   ::  ::    : :: :::::::::: ::: ::    :  :: ::::::
USPN   CCGGCTGGCGTGCGGAGACACTCGGAGACGACATCGCATGGATGCGATTTGGCAAGGACG
```

```
        860        870        880        890        900        910

    1020          1030       1040       1050       1060       1070
./wwwt G---CCTCTACGCAGTTAACCCAGAAAATGGTTTCTTCGGTGTTGCTCCAGGCACCAACT
       :   ::: ::::: :: ::::: :::    :: :::::::: :: :: :: ::::::::::
USPN   GTCGCCTGTACGCCGTCAACCCGGAATTCGGCTTCTTCGGGGTGGCGCCGGGCACCAACT
        920        930        940        950        960        970

       1080       1090       1100       1110       1120       1130
./wwwt ACGCATCCAACCCAATCGCGATGAAGACCATGGAACCAGGCAACACCCTGTTCACCAACG
          :: ::::: : ::: :::   ::::: :   : :::::::  :::::::::: :
USPN   GGAAGTCGAACCCTAACGCCATGCGCACCATTGCCGCCGGCAACACGGTGTTCACCAATG
        980        990       1000       1010       1020       1030

       1140       1150       1160       1170       1180       1190
./wwwt TGGCACTCACCGACGACGGCGACATCTGGTGGGAAGGCATGGACGGCGACGCCCCAGCTC
       : ::::::::::::::::::::: : :::::::: ::: :::: :::::: :: ::::
USPN   TCGCACTCACCGACGACGGCGACGTGTGGTGGGAGGGCCTGGAAGGCGAC-CCGCAGC--
       1040       1050       1060       1070       1080       1090

       1200       1210       1220          1230       1240       1250
./wwwt ACCTCATTGACTGGATGGGCAACGACTGG-AC--CCCAGAGTCCGACGAAAACGCTGCTC
       :::: :: ::::::: :::::::::::::: ::  ::  ::: : ::    :: :: :: :
USPN   ACCTGATCGACTGGAAGGGCAACGACTGGTACTTCCGCGAGACGGAAACCAATGCGGCAC
          1100       1110       1120       1130       1140       1150

          1260       1270       1280       1290       1300       1310
./wwwt ACCCTAACTCCCGTTACTGCGTAGCAATCGACCAGTCCCCAGCAGCAGCACCTGAGTTCA
       :::: :::::::: ::::::  : : ::    :::: :::        :: :: ::::
USPN   ACCCGAACTCCCGGTACTGCACACCGATGTCGCAGTGCCCGATCCTGGCCCCCGAGTGGG
          1160       1170       1180       1190       1200       1210

          1320       1330       1340       1350       1360        1370
./wwwt ACGACTGGGAAGGCGTCAAGATCGACGCAATCCTCTTCGGTGGACGTCGCGCAGAC-ACC
       : :::  : : ::::::  ::::   :  ::::: ::::: :: :: :::  :::: ::
USPN   ATGACCCGCAGGGCGTCCCGATCTCGGGGATCCTGTTCGGCGGCCGCCGCA-AGACCACG
          1220       1230       1240       1250       1260        1270

           1380       1390       1400       1410       1420       1430
./wwwt GTCCCACTGGTTACCCAGACCTACGACTGGGAGCACGGCACCATGGTTGGTGCACTGCTC
       :: :: ::::: ::: :: :   ::::::: :::::::     :  : :::::    ::
USPN   GTTCCGCTGGTCACCGAGGCGCGCGACTGGCAGCACGGGGTGTTCATCGGTGCGACCCTG
           1280       1290       1300       1310       1320       1330

           1440       1450       1460       1470       1480       1490
./wwwt GCATCCGGTCAGACCGCAGCTTCCGCAGAAGCAAAGGTCGGCACACTCCGCCACGACCCA
       :    ::  :::::::: ::  :::   : :  ::::::::::   : :::: ::::::
USPN   GGTAGCGAGCAGACCGCCGCGGCCG---AGGGCAAGGTCGGCAATGTGCGCCGCGACCCG
           1340       1350          1360       1370       1380

           1500       1510       1520       1530       1540       1550
./wwwt ATGGCAATGCTCCCATTCATTGGCTACAACGCTGGTGAATACCTGCAGAACTGGATTGAC
       :::::: ::::: :: ::  : :::::::::: ::: :: ::: : ::: :::::::  ::
USPN   ATGGCCATGCTGCCGTTTTTGGGCTACAACGTTGGGGACTACTTCCAGCACTGGATCAAC
    1390       1400       1410       1420       1430       1440

           1560       1570          1580       1590       1600
./wwwt ATGGGTAACAAGGGTGGCGA---CAAGATGCCATCCATCTTCCTGGTCAACTGGTTCCGC
        :::: ::  : :  :  ::   :::: ::::      : ::: : :::::::::::::
USPN   CTGGGCAAGCACGCCGATGAGTCCAAGCTGCCCAAGGTGTTCTTCGTCAACTGGTTCCGT
```

```
    1450       1460       1470       1480       1490       1500

    1610       1620       1630       1640       1650       1660
./wwwt CGTGGCGAAGATGGACGCTTCCTGTGGCCTGGCTTCGGCGACAACTCTCGCGTTCTGAAG
       :: :: :: :: :: :::::::::::::: :::::::::: ::: :: :: ::::::
USPN   CGCGGTGACGACGGTCGCTTCCTGTGGCCGGGCTTCGGCGAGAACAGCCGGGTGCTGAAG
    1510       1520       1530       1540       1550       1560

    1670       1680       1690       1700       1710       1720
./wwwt TGGGTCATCGACCGCATCGAAGGCCACGTTGGCGCAGACGAG-ACCGTTGTTGGACACAC
       ::: :: :::: ::::::::   : : :  :::: : ::: ::: : ::  :::
USPN   TGGATCGTCGATCGCATCGAGCACAAGGCCGGCGGTG-CGACCACCCCGATCGG---CAC
    1570       1580       1590       1600       1610       1620

     1730       1740       1750       1760       1770       1780
./wwwt CGCTAAGGCCG---AAGACCTCGACCTCGACGGCCTCGACACCCCAATTGAGGATGTCAA
       :: :   ::::   : ::: : ::::: ::::: :: ::: :       :  :::::
USPN   CGTTCCCGCCGTGGAGGACTTGGACCTGGACGGACTGGACGTCGACGCCGCCGATGTAGC
        1630       1640       1650       1660       1670       1680

        1790       1800       1810       1820       1830       1840
./wwwt GGAAGCACTGACCGCTCCTGCAGAGCAGTGGGCAAACGACGTTGAAGACAACGCCGAGTA
        :  :: ::: : :    ::: ::  : :::    : ::  :        : ::  :: :
USPN   CGCGGCGCTGGCAGTCGATGCCGATGAATGGCGTCAGGAACTGCCGCTGATCGAAGAATG
        1690       1700       1710       1720       1730       1740

        1850       1860       1870        1880        1890       1900
./wwwt CCTCACTTTCCTCGGACCACGTGTTCCTGCAGA--GGT-TCACAGCC-AGTTCGATGCTC
        ::    ::: ::::    :: :   :::: ::  ::: ::: ::   ::::::: :: :
USPN   GCTGCAGTTCGTCGG----CGAGAAGCTGCCGACCGGTGTCAAAGATGAGTTCGACGCCC
        1750       1760           1770       1780       1790       1800

            1910       1920       1930       1940       1950
./wwwt TGAAGGCCCGCATTTCAGCAGCTCACGCTTAAAGTTCACGCTTAAGAACTGC
       ::::::             :::    :: ::
USPN   TGAAGG-------------AGC----GCCTA--------------------
                          1810
```

Back to ISREC bioinformatics group home page

APPENDIX L

NiceSite View of: PS00505

General information about the entry	
Entry name	**PEPCK_GTP**
Accession number	**PS00505**
Entry type	PATTERN
Date	MAY-1991 (CREATED); APR-2006 (DATA UPDATE); MAY-2006 (INFO UPDATE).
PROSITE documentation	PDOC00421
Name and characterization of the entry	
Description	Phosphoenolpyruvate carboxykinase (GTP) signature.
Pattern	[FY]-P-S-[AGMS]-C-G-K-T-[NS].

Numerical results

- UniProtKB/Swiss-Prot release number: **50.1**, total number of sequence entries in that release: **223100**.
- Total number of hits in UniProtKB/Swiss-Prot: **46 hits in 46 different sequences**
- Number of hits on proteins that are known to belong to the set under consideration: **46 hits in 46 different sequences**
- Number of hits on proteins that could potentially belong to the set under consideration: **0 hits in 0 different sequences**
- Number of false hits (on unrelated proteins): **0 hits in 0 different sequences**
- Number of known missed hits: **5**
- Number of partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences: **1**
- Precision (true hits / (true hits + false positives)): **100.00 %**
- Recall (true hits / (true hits + false negatives)): **90.20 %**

Comments

- Taxonomic range: **Archaebacteria, Eukaryotes, Prokaryotes (Bacteria)**

- Maximum known number of repetitions of the pattern in a single protein: **1**
- `Interesting' site in the pattern: **5,active_site(?)**
- VERSION: **2**

Cross-references

UniProtKB/Swiss-Prot

True positive hits:

PPCKC_BOVIN (Q8HYZ4), PPCKC_CHICK (P05153), PPCKC_HUMAN (P35558),
PPCKC_MOUSE (Q9Z2V4), PPCKC_PONPY (Q5R5J1), PPCKC_RAT (P07379),
PPCKM_CHICK (P21642), PPCKM_HUMAN (Q16822), PPCKM_MOUSE (Q8BH04),
PPCK_ACIAD (Q6F8P2), PPCK_ASCSU (Q05893), PPCK_AZOSE (Q5P2P8),
PPCK_BORBR (Q7WJQ9), PPCK_BORPA (Q7WAK8), PPCK_BURMA (Q62FI7),
PPCK_BURPS (Q63VB7), PPCK_CHLAB (Q5L4X1), PPCK_CHLCV (Q821M4),
PPCK_CHLLI (Q08262), PPCK_CHLMU (Q9PLL6), PPCK_CHLPN (Q9Z755),
PPCK_CHLTE (Q8KAD1), PPCK_CHLTR (O84716), PPCK_CORDI (Q6NET5),
PPCK_COREF (Q8FM16), PPCK_CORGL (Q9AEM1), PPCK_CORJK (Q4JY04),
PPCK_DROME (P20007), PPCK_GEOSL (Q746Y3), PPCK_HAECO (P29190),
PPCK_LEIXX (Q6AGS4), PPCK_MYCBO (P65687), PPCK_MYCLE (O06084),
PPCK_MYCPA (Q73TS2), PPCK_MYCSM (Q9AGJ6), PPCK_MYCTU (P65686),
PPCK_NEOFR (P22130), PPCK_NOCFA (Q5YNB0), PPCK_PYRAB (Q9UY53),
PPCK_PYRFU (Q8U410), PPCK_PYRHO (O58050), PPCK_PYRKO (Q6F494),
PPCK_RALSO (Q8Y3G3), PPCK_STRAW (Q82I71), PPCK_STRCO (Q93JL5),
PPCK_TREPA (O83159)

False negative hits (sequences which belong to the set under consideration, but which have not been picked up by the pattern or profile):

PPCK_SULAC (Q4J9S8), PPCK_SULSO (Q97VS5), PPCK_SULTO (Q972S7),
PPCK_THEAC (Q9HLV2), PPCK_THEVO (P58306)

`Potential' hits (partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences):

PPCK_FASHE (P80525)

	Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits: [Clustal format, color, condensed view] [Clustal format, color] [Clustal format, plain text] [Fasta format]
PDB [Detailed view]	1KHB; 1KHE; 1KHF; 1KHG; 1M51; 1NHX;

View entry in original PROSITE format View entry in raw text format (no links) Direct ScanProsite submission

---

If you would like to retrieve all the Swiss-Prot entries referenced in the DR lines of this entry (with the exception of false positive hits) , you can enter a file name. These entries will then be saved to a file under this name in the directory `outgoing` of the ExPASy anonymous ftp server, from where you can download it. (Please note that this temporary file will only be kept for 1 week.)

File name:

Format: ◉ Swiss-Prot ○ Fasta

Reset or Create file

Documentation PDOC00421

Phosphoenolpyruvate carboxykinase (GTP) signature

Description:

Phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) (PEPCK) [1] catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate while hydrolyzing GTP, a rate limiting step in gluconeogenesis (the biosynthesis of glucose). In vertebrates there are two isozymes: a cytosolic form whose activity is affected by hormones regulating this metabolic process (such as glucagon, or insulin) and a mitochondrial form.

An essential cysteine residue has been proposed [2] to be implicated in the catalytic mechanism; this residue is located in the central part of PEPCK and is in the center of a perfectly conserved region that we use as a signature pattern.

Note:
Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) an enzyme that catalyzes the same reaction, but using ATP instead of GTP, is not related to the above enzyme (see <PDOC00460>).

Last update:
April 2006 / Pattern and text revised.

Technical section:

PROSITE method (with tools and information) covered by this documentation:

PEPCK_GTP, PS00505; **Phosphoenolpyruvate carboxykinase (GTP) signature** (PATTERN)

[FY] - P - S - [AGMS] - C - G - K - T - [NS]

Consensus pattern:	*C may be the active site residue*
Sequences known to belong to this class detected by the pattern:	ALL
Other sequence(s) detected in Swiss-Prot:	NONE.

- Retrieve an alignment of Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Taxonomic tree view of all Swiss-Prot/TrEMBL entries matching PS00505
- Retrieve a list of all Swiss-Prot/TrEMBL entries matching PS00505
- Scan Swiss-Prot/TrEMBL entries against PS00505
- view ligand binding statistics

Matching PDB structures: 1KHB 1KHE 1KHF 1KHG ... [ALL]

References:

1	*Authors*	Weldon S.L., Rando A., Matathias A.S., Hod Y., Kalonick P.A., Savon S., Cook J.S., Hanson R.W.
	Title	*Mitochondrial phosphoenolpyruvate carboxykinase from the chicken. Comparison of the cDNA and protein sequences with the cytosolic isozyme.*
	Source	J. Biol. Chem. 265:7308-7317(1990).
	PubMed ID	2110163
2	*Authors*	Lewis C.T., Seyer J.M., Carlson G.M.
	Title	*Cysteine 288: an essential hyperreactive thiol of cytosolic phosphoenolpyruvate carboxykinase (GTP).*
	Source	J. Biol. Chem. 264:27-33(1989).
	PubMed ID	2909519

Copyright:

Miscellaneous:

View entry in original PROSITE document format
View entry in raw text format (no links)

APPENDIX M

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 264, No. 1, Issue of January 5, pp. 27-33, 1989
Printed in U.S.A.

Cysteine 288: An Essential Hyperreactive Thiol of Cytosolic Phosphoenolpyruvate Carboxykinase (GTP)*

(Received for publication, June 23, 1988)

Cristina T. Lewis‡, Jerome M. Seyer§, and Gerald M. Carlson¶

From the Department of Biochemistry, College of Medicine, The University of Tennessee, Memphis, Tennessee 38163

Phosphoenolpyruvate carboxykinase from the cytosol of rat liver has 13 cysteines, at least one of which is known to be very reactive and essential for catalytic activity (Carlson, G. M., Colombo, G., and Lardy, H. A. (1978) *Biochemistry* 17, 5329–5338). In order to identify the essential cysteine, this enzyme was modified with the fluorescent sulfhydryl reagent *N*-(7-dimethylamino-4-methyl-3-coumarinyl)maleimide. Incubation of phosphoenolpyruvate carboxykinase with a 10% molar excess of this maleimide at 0 °C results in the rapid and nearly complete loss of catalytic activity. Under these conditions, 1 mol of the maleimide is incorporated per mol of inactivated enzyme. The substrate GDP provides almost complete protection against inactivation and modification, while phosphoenolpyruvate protects against the rate, but not the extent, of modification. The pH dependence of the rate of enzyme inactivation suggests that the modified residue has a pK_a of approximately 7.0. Purification and sequencing of the labeled peptide identifies the hyperreactive essential cysteine as Cys-288. This cysteine lies between two putative phosphoryl-binding domains and within a hydrophobic sequence.

Phosphoenolpyruvate carboxykinase (EC 4.1.1.32) catalyzes the conversion of oxaloacetate to phosphoenolpyruvate in the first committed step of gluconeogenesis. Both cytosolic and mitochondrial isoenzymes of phosphoenolpyruvate carboxykinase exist, and the relative proportions of each vary among different species. Comparison of a common region of the cDNA sequences of the avian isoenzymes indicates that they share 63% homology at the amino acid level within that particular segment (1). In general, the isoenzymes appear to have similar kinetic properties but can be distinguished immunologically, electrophoretically, and chromatographically (2–7). Recently, a great deal of attention has been focused on the gene for the cytosolic carboxykinase and its transcriptional regulation by cAMP, insulin, and glucocorticoids. The structures of the gene and messenger RNA for the cytosolic isoenzyme from rat liver have been published, and the amino acid sequence has been deduced from the cDNA sequence and confirmed by fast atom bombardment mass spectrometry (8). However, very little is known about the structure of the enzyme's active site or about the amino acids that may participate in catalysis.

Phosphoenolpyruvate carboxykinase that has been isolated from sources as disparate as oyster and pig liver is sensitive to sulfhydryl reagents, and the sulfhydryl chemistry of the enzyme has been studied in a number of laboratories. The cytosolic enzyme from rat liver has 13 cysteines, all in the reduced state, among its 622 amino acid residues (M_r = 69,289)[1] (8, 9). Low concentrations of exogenous thiols stabilize or optimize enzyme activity (2, 10–14), while disulfides promote loss of activity (15, 16). The cytosolic carboxykinase from rat liver (17) and the mitochondrial enzyme from sheep kidney (18–20), chicken liver (21), and bullfrog liver (22) each contain at least one cysteine that is essential for catalytic activity. Furthermore, the inactivation of the enzyme by a variety of sulfhydryl reagents is substantially diminished in the presence of substrates; in each case nucleotide or a combination of nucleotide and Mn^{2+} provides the most effective protection against inactivation (17, 19, 21, 22). In no case, however, has the identity or function of a critical cysteine of phosphoenolpyruvate carboxykinase been reported.

Our studies have focused on the cytosolic carboxykinase from rat liver, which is known to contain a hyperreactive essential cysteine that lies proximal to a second cysteine (17). Incubation of this enzyme with equimolar Nbs_2[2] in the presence of phosphoenolpyruvate resulted in the rapid and nearly complete loss of activity and the release of 1 mol of thionitrobenzoate/mol of enzyme. Plots of residual activity against moles of cysteine modified per mol of enzyme extrapolated to one at zero activity. This critical cysteine is near a second cysteine because incubation with equimolar Nbs_2 in the absence of substrates resulted in the formation of an intramolecular cystine disulfide linkage that included an essential cysteine (2.0 mol of thionitrobenzoate released per mol of enzyme and a residual activity of 3%). Inactivation of the enzyme by reagents with preferential reactivity toward dithiols provided additional evidence for this vicinal dithiol grouping. The critical cysteine appears to be truly essential because the ability of the modified enzyme to catalyze the formation of phosphoenolpyruvate, oxaloacetate, or pyruvate was nearly completely destroyed, and replacement of the

* This work was supported in part by Research Grant DMB 85-20311 from the National Science Foundation (to G. M. C.) and by research funds from the Veterans Administration (to J. M. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of the Doggett Predoctoral Fellowship from the College of Graduate Health Sciences, The University of Tennessee, Memphis.

§ Primary address: Veterans Administration Medical Center, 1030 Jefferson Ave., Memphis, TN 38104.

¶ To whom correspondence should be sent.

[1] The residue numbering system used herein is based upon 622 amino acids and includes the amino-terminal methionine as residue 1. It is not clear if the NH_2-terminal methionine is cleaved or modified *in vivo*, because its presence has not been confirmed by amino-terminal sequencing (8).

[2] The abbreviations used are: Nbs_2, 5,5′-dithiobis(2-nitrobenzoate); DACM, *N*-(7-dimethylamino-4-methylcoumarinyl)maleimide; Hepes, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; HPLC, high-performance liquid chromatography; MES, 2-(N-morpholino)ethanesulfonic acid; TES, *N*-tris[hydroxymethyl]methyl-2-aminoethanesulfonic acid.

bulky nitrobenzoate moiety with cyanide did not result in any reactivation. The essential cysteine is surprisingly reactive; formation of the intramolecular disulfide occurred approximately 25 times faster than the analogous reaction of Nbs_2 with dithiothreitol. Although the critical cysteine was not identified in that study, specific cleavage at the ^{14}C-cyanylated cysteine indicated that the residue in question is approximately 44% of the distance from the amino terminus along the peptide backbone (17). Three cysteines lie within this region of the enzyme: Cys-244 at 39%, Cys-288 at 46%, and Cys-306 at 49% of the distance from the amino terminus (8). In this study we report the specific modification and inactivation of phosphoenolpyruvate carboxykinase by a fluorescent sulfhydryl reagent and the identification of the critical modified thiol as cysteine 288.

EXPERIMENTAL PROCEDURES

Reagents—DACM was from Serva Fine Biochemicals, *N*-ethylmaleimide was from Sigma, and the remaining maleimides were from Molecular Probes, Inc. DACM was stored in acetone at −20 °C; its concentration was determined spectrophotometrically after dilution into buffer and filtration. The molar extinction coefficients used for the maleimide derivatives were: DACM, 22,600 at 395 nm (23); 7-diethylamino-3-(4′-maleimidylphenyl)-4-methylcoumarin, 30,200 at 387 nm; and *N*-(1-pyrene)maleimide, 37,500 at 343 nm (24). TES, MES, Tris, Hepes, dithiothreitol, and ultrapure urea were from Research Organics; phosphoenolpyruvate, oxaloacetate, NADH, and Nbs_2 were from Boehringer Mannheim. Agarose-hexane-GTP and all nucleotides were from Pharmacia LKB Biotechnology, Inc. HPLC grade ammonium acetate and ultrapure sodium phosphate were from J. T. Baker Chemical Co. and Aldrich, respectively.

Enzymes—Glycerol solutions of lactate dehydrogenase, malate dehydrogenase, and pyruvate kinase were from Boehringer Mannheim. Trypsin (treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone) was from Sigma.

Phosphoenolpyruvate carboxykinase was purified to homogeneity from rat liver cytosol using published procedures (9) with the following modifications. Frozen livers from male 24-h fasted rats (Cocalico Biologicals, Inc.) were thawed for 45 min in cold buffer prior to homogenization. The ammonium sulfate back extraction was omitted from the purification procedure. The final purification of the carboxykinase by chromatography on agarose-hexane-GTP provided homogeneous enzyme with a high constant specific activity of 19–23 μmol of oxaloacetate formed/min/mg of protein at 25 °C. Exogenous thiols were removed by gel filtration in the presence of 10 mM TES (pH 6.9), 5% glycerol, and 0.5 mM EDTA, and the enzyme was stored at 5 °C. Phosphoenolpyruvate carboxykinase concentration was determined spectrophotometrically using a molar extinction coefficient[3] of 1.15×10^5; the concentrations in this report refer to enzyme molarity and not to total cysteine molarity. All buffers were degassed and saturated with nitrogen.

Activity Assays—The rates of formation of phosphoenolpyruvate and oxaloacetate were determined spectrophotometrically as previously described (9) except that 100 μM EDTA and 200 μM $MnCl_2$, in addition to 4 mM $Mg(CH_3CO_2)_2$, were included in the assay measuring formation of phosphoenolpyruvate. The rate of formation of pyruvate from oxaloacetate was measured as described (25) in the presence of 56 mM Hepes (pH 7.0), 1.0 mM IDP, 250 μM NADH, 40 μg of lactate dehydrogenase, 600 μM oxaloacetate, 250 μM EDTA, and 350 μM $MnCl_2$. All spectrophotometric analyses were performed on a Beckman DU-70 or Gilford Response spectrophotometer, and all activity assays were carried out at 25 °C.

Modification and Determination of Stoichiometry of Labeling—Under the standard conditions of modification, phosphoenolpyruvate carboxykinase (3–10 μM) was incubated with a 1.1-fold molar excess of DACM at 0 °C in the presence of 10 mM TES (pH 6.9), 5% glycerol, and 0.5 mM EDTA. At appropriate intervals aliquots were diluted 100-fold and assayed for enzyme activity.

For determination of the stoichiometry of labeling, the modification was terminated after 50 min by the addition of dithiothreitol (to 2 mM) or β-mercaptoethanol (to 1%, v/v); urea was then added to 6 M. In the presence of phosphoenolpyruvate, modifications for determination of stoichiometry were terminated at an average time of 15 min in order to achieve approximately 50% inactivation. Quantitation of labeling by DACM required conversion of the reaction product into a more stable fluorescent derivative, which was accomplished by incubation at alkaline pH (23). The reaction mixture was either dialyzed against 50 mM ammonium bicarbonate (pH 9.0), 0.01% (v/v) β-mercaptoethanol, and 6 M urea for 24 h at 25 °C or incubated at pH 9.0 in the presence of 6 M urea and subsequently desalted over Sephadex G-50 equilibrated in 50 mM ammonium bicarbonate, pH 8.5, 0.1% (v/v) β-mercaptoethanol, and 6 M urea. Protein concentration was determined using the Bio-Rad protein assay with the standard curve constructed using unmodified phosphoenolpyruvate carboxykinase. DACM concentration was determined spectrophotometrically using an extinction coefficient of 19,800 M^{-1} cm^{-1} at 383 nm for the stable DACM-protein adduct (23).

Total thiol titrations with Nbs_2 of the native and modified enzyme were performed by bringing the reaction mixtures to 0.2% sodium dodecyl sulfate or to 6 M urea; reactions were initiated with excess Nbs_2. The absorbance at 412 nm of each sample was monitored with time at 25 °C, from which was subtracted a blank that contained all components of the reaction mixture except enzyme.

Determination of the Apparent pK_a—The pH dependence of enzyme inactivation by DACM was measured by incubating phosphoenolpyruvate carboxykinase (1 μM) with equimolar DACM at 0 °C in the presence of 20 mM each of acetic acid/MES/TES/Tris (at the appropriate pH), 5% glycerol, and 78 μM EDTA. During the 30-min incubation, aliquots of the reaction mixture were diluted 100-fold, and the rate of formation of oxaloacetate was measured. All modifications were initiated with enzyme, and the results were assessed relative to a control in which the enzyme was incubated at the same pH, but in the absence of DACM. The effect of pH on DACM integrity was assessed by preincubating DACM for 1 h at the appropriate pH before adjusting the pH to 7 and measuring the probe's absorption spectrum and its ability to subsequently inactivate the carboxykinase at pH 7.

The apparent pK_a of the modified residue was also estimated by measuring the effect of pH on the second order rate constant for the reaction of *N*-ethylmaleimide with the enzyme. This reaction was initiated with enzyme (1 μM) and contained 20 mM each acetic acid/MES/TES/Tris (at the appropriate pH), 5% glycerol, 0.5 mM EDTA, and a 2-fold molar excess of *N*-ethylmaleimide over enzyme. During the 10-min incubation at 25 °C, aliquots were diluted 100-fold, and the rate of oxaloacetate formation was measured; results were assessed relative to the control enzyme that was incubated at the same pH in the absence of *N*-ethylmaleimide. The second order rate constants for the inactivation were calculated from the slopes of plots of $\ln(A - P)/(B - P)$ *versus* time, where A and B are the initial concentrations of *N*-ethylmaleimide and phosphoenolpyruvate carboxykinase, respectively, and P is the concentration of inactivated enzyme ($B \times$ fractional activity loss at time t). The slope of such a plot is equal to $k(A - B)$, where k is the observed second order rate constant. The effect of pH on the rate of inactivation was assessed according to the following equation (26),

$$1/k = [1/(k_{max} \cdot K)] \cdot [H^+] + 1/k_{max} \quad (1)$$

where k is the observed second order rate constant at a given pH, k_{max} is the maximal rate constant, and K is the ionization constant of the modified residue. Thus, the ratio of the slope to the intercept of a plot of $1/k$ against $[H^+]$ is equal to $1/K$.

The pHs tested with these reagents ranged from 4 to 9 for DACM and 5.5 to 8 for *N*-ethylmaleimide. There was no significant loss in the activity of control enzyme incubated in the absence of sulfhydryl reagents at pH 6–8, and there was a maximal 30% loss in activity when incubated at pH 4. At pH 9 and above, DACM was partially destroyed, as evidenced by a shift in its absorption spectrum and a greater than 50% decrease in its ability to cause inactivation at pH 7 when it was first preincubated at pH 9. Below pH 6, the extent of inactivation of phosphoenolpyruvate carboxykinase by either *N*-ethylmaleimide or DACM was low. Therefore, to calculate the apparent pK_a, we restricted our analyses to the data obtained at pH 6–8.

Tryptic Digestion and Reversed-phase HPLC—The modified denatured enzyme was carboxymethylated with iodoacetate (27) and dialyzed against 50 mM ammonium bicarbonate (pH 9.0) and 2 M urea for 24 h at 25 °C. The pH was adjusted to 8.0, and the sample was digested with trypsin (2% by weight) for 12 h at 37 °C, with an

[3] The molar extinction coefficient was based on a previous determination and was corrected for the exact molecular weight of phosphoenolpyruvate carboxykinase (8, 9).

additional 2% by weight of trypsin added after 6 h.

Purification of tryptic peptides by HPLC was carried out with a Waters HPLC system equipped with a model 490 multiwavelength detector. The lyophilized tryptic digest was dissolved in 10 mM sodium phosphate (pH 6.5) and subjected to reversed-phase HPLC on a Radial-Pak μBondapak C_{18} column (8 mm × 10 cm, Waters Associates) at a flow rate of 1 ml/min. Peptides were eluted with a triphasic linear gradient of acetonitrile: 0–25% from 5 to 45 min, 25–35% from 45 to 110 min, and 35–80% from 110 to 140 min. Fractions of 1 ml were collected, and the modified peptide was detected and pooled, based upon its absorbance at 210 and 383 nm. The partially purified peptide was lyophilized and dissolved in 10 mM ammonium acetate (pH 6.0), 4 M guanidine HCl, and applied to the same column in 10 mM ammonium acetate (pH 6.0) at a flow rate of 1.0 ml/min. The purified peptide was eluted with a triphasic linear gradient of acetonitrile: 0–25% from 5 to 45 min, 25–35% from 45 to 145 min, and 35–80% from 145 to 175 min. Fractions of 1 ml were again collected, and the purified peptide was lyophilized prior to sequence analysis.

Amino Acid Sequencing—Automated Edman degradations were performed with an automatic sequenator (Applied Biosystems model 477 pulse-liquid sequenator) according to published procedures (28). The amino acids were converted using the automated trifluoroacetic acid conversion program, and the released phenylthiohydantoin derivatives were identified and quantified using a Waters HPLC equipped with a C_{18} μBondapak column (29).

RESULTS

Inactivation of Phosphoenolpyruvate Carboxykinase by DACM—Previous studies have shown that modification by Nbs_2 of a single cysteine of phosphoenolpyruvate carboxykinase from the cytosol of rat liver results in the almost complete loss of catalytic activity (17). Our first goal was to establish conditions whereby a single critical cysteine of the enzyme could be covalently and irreversibly labeled in a highly specific manner. Incubation of the carboxykinase with only a 10% molar excess of the fluorescent maleimide derivative DACM at 0 °C resulted in the very rapid and nearly complete loss of catalytic activity (Fig. 1). In some instances the rate of inactivation was slightly slower, but in all cases the time required for 50% inactivation was less than 2 min. The three activities of phosphoenolpyruvate carboxykinase, namely the formation of phosphoenolpyruvate, oxaloacetate, and pyruvate, were lost in parallel, in agreement with previous findings obtained with Nbs_2 (17). The residual activity was typically less than 10%, but only once did we observe a residual activity less than 1%. The second order rate constant for the reaction was not determined, because the rapidity of the reaction precluded the generation of linear second order plots of the data and because the ratio of inhibitor to enzyme was necessarily low.

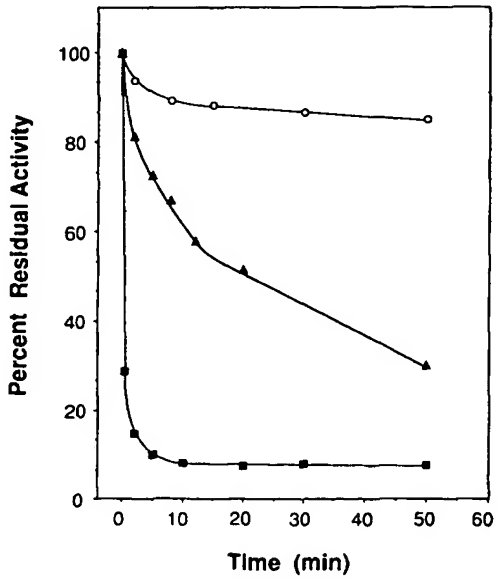

FIG. 1. **Inactivation of phosphoenolpyruvate carboxykinase by DACM and substrate protection.** Enzyme (1 μM) was incubated with a 1.1-fold molar excess of DACM at 0 °C in the absence (■) or presence of 5 mM phosphoenolpyruvate (▲) or 1 mM GDP (○). The rate of formation of oxaloacetate was measured at the indicated times as described under "Experimental Procedures."

Previous studies with the mitochondrial carboxykinase have provided evidence for an essential hydrophobic domain (30, 31). *N*-(Iodoacetylaminoethyl)-5-naphthylamine-1-sulfonate behaves as an affinity label for phosphoenolpyruvate carboxykinase in that it exhibited saturation kinetics of inactivation and specific substrate protection. In addition, 1-anilinonaphthalene-8-sulfonate as well as several other anilinonaphthalene sulfonate isomers also inactivated the enzyme, but without apparent covalent modification (30, 32). The initial rapid inactivation of the carboxykinase by DACM, however, is not merely due to tight noncovalent binding of the probe, because incubation of the enzyme with a DACM-mercaptoethanol conjugate had no effect on catalytic activity. Furthermore, there was no significant difference in the inactivation profiles if aliquots of a reaction mixture were assayed immediately or if they were first diluted 20-fold in the presence of dithiothreitol and then assayed after an additional 10-min incubation (data not shown).

Abilities of Substrates to Protect against Inactivation—Fig. 1 and Table I show that the pattern of substrate protection against inactivation by DACM is similar to that described in earlier reports with other sulfhydryl reagents (17–20, 22). Nucleotides provided almost complete protection against inactivation, while phosphoenolpyruvate protected against the rate but not the extent of modification. Oxaloacetate also affected only the rate of inactivation, but it was less effective than phosphoenolpyruvate; the rates of the loss in phosphoenolpyruvate- and pyruvate-forming activity were similar in the presence of 1 mM oxaloacetate. The tri-, di-, and monophosphates of guanosine nucleotides afforded similar protection. Metal nucleotides were slightly more effective than free nucleotides, but neither free Mg^{2+} nor free Mn^{2+} had any effect on the rate of inactivation. Thus, the loss in catalytic activity upon modification of the sulfhydryl does not appear to be due to a perturbation of the activating divalent cation binding

TABLE I

Substrate protection against inactivation by DACM

Phosphoenolpyruvate carboxykinase was incubated with a 1.1-fold molar excess of DACM at 0 °C, and the inactivation was determined as described under "Experimental Procedures." The percent residual activity after 30 min is shown for a typical experiment. All assays measured oxaloacetate formation unless otherwise indicated.

Addition	Residual activity
	%
None	7.6
1 mM GDP	94.6
1 mM GTP	89.3
1 mM GMP	80.0
1 mM MgGDP	94.0
1 mM MnGDP	95.7
1.5 mM Mg^{2+}	5.5
1.5 mM Mn^{2+}	4.8
5 mM Phosphoenolpyruvate	24.7
1 mM Oxaloacetate	12.4[a]
1 mM Phosphoenolpyruvate + GDP	95.6[a]
1 mM Mn-GDP-oxaloacetate	94.1[a]
1 mM Mn-GDP-oxaloacetate	96.4[b]

[a] The loss in phosphoenolpyruvate-forming activity was measured.
[b] The loss in pyruvate-forming activity was measured.

site. Ternary complexes were neither more nor less effective than nucleotides alone in protecting against inactivation.

Stoichiometry of Labeling—If a residue is essential for catalytic activity, then the stoichiometry of labeling of that residue should correlate with the loss of enzyme activity. For five experiments in which the incorporation of DACM into phosphoenolpyruvate carboxykinase was measured after the inactivation had reached a plateau, a value of 1.11 ± 0.08 (mean ± S.E.) mol of DACM/mol of inactivated enzyme was obtained. Further evidence that a critical cysteine was specifically labeled was obtained when the enzyme was modified in the presence of substrates. There was no detectable incorporation of DACM when 1 mM GDP was included in a 50-min incubation under standard conditions, which is consistent with the very low levels of inactivation (Table I). For modifications in the presence of 5 mM phosphoenolpyruvate, the reactions were purposely terminated at about 50% residual activity; under these conditions an average of 0.99 ± 0.09 mol of DACM was incorporated per mol of inactivated enzyme (mean ± S.E., $n = 5$). In summary, the incorporation of the fluorescent probe correlated very well with enzyme inactivation, regardless of whether the loss in activity was minimal, intermediate, or nearly complete. We wish to emphasize the importance of using chemically homogeneous enzyme for such experiments. Phosphoenolpyruvate carboxykinase that has been purified by chromatography on agarose-hexane-GTP has been freed not only from contaminating proteins but also from a low specific activity form of the enzyme (9). Comparison of total thiol titrations (with Nbs_2) of the native and modified enzyme confirmed that the modified residue was indeed a cysteine, but we were unable to precisely quantitate the moles of residual cysteine using this method because it required detection of a small change in A_{412} in the presence of a large background absorbance by DACM.

Effect of Hydrophobic Maleimide Derivatives—As previously discussed, the amino acid sequence of phosphoenolpyruvate carboxykinase indicates that there are three candidates for the essential cysteine, based upon their distance from the amino terminus. It has been suggested that Cys-245 may be the best candidate for the critical thiol, inasmuch as Cys-288 is located within a hydrophobic region of the sequence and might therefore be buried within the interior of the molecule (8). We attempted to characterize the microenvironment of the critical cysteine by comparing the abilities of different maleimide derivatives to inactivate the enzyme. The chemical structures of the maleimide derivatives are shown in Fig. 2*A*, and their relative potencies for inactivation are illustrated in Fig. 2*B*. *N*-Ethylmaleimide was clearly much less effective than any of the hydrophobic maleimides, and *N*-(1-pyrene)maleimide, the most hydrophobic compound, was in fact the most potent. While it is true that the more bulky derivatives will be more likely to perturb the overall structure of the enzyme and therefore cause a greater degree of inactivation, the efficacy of the derivatives followed more closely a trend in increasing hydrophobicity than in increasing size. In any case, even the bulky hydrophobic compounds were not excluded from modifying a critical thiol, and the results are consistent with the cysteine existing in a hydrophobic microenvironment.

Determination of an Apparent pK_a—Fig. 3*A* shows the effect of pH on the extent of inactivation of phosphoenolpyruvate carboxykinase by equimolar DACM at a fixed time (*closed symbols*) and on the rate of inactivation by *N*-ethylmaleimide (*open symbols*). The *closed symbols* represent the extent of inactivation after 5 min under conditions in which the loss in activity had not plateaued. The rate of inactivation by DACM under these conditions was somewhat slower than usual (*cf.* Fig. 1), perhaps due to a difference in ionic strength of the buffer or to the slightly lower DACM concentration (equimolar *versus* 10% excess of equimolar). The titration curves are similar and have inflection points of approximately 6.8 (DACM curve) and 7.0 (*N*-ethylmaleimide curve). Because we were unable to determine the second order rate constants for the reaction of phosphoenolpyruvate carboxykinase with DACM, we instead modified the enzyme with *N*-ethylmaleimide, a reagent that inactivates the enzyme more slowly than does DACM (see Fig. 2*B*) and thus permits determination of second order kinetic constants. The rate constants for this reaction were determined at various pHs as described under "Experimental Procedures," and the reciprocals of the observed rate constants were plotted against $[H^+]$ (Fig. 3*B*). From the slope and intercept of this plot the apparent pK_a of the modified residue was calculated to be 7.0. The pH dependence illustrated in Fig. 3 is not due to an effect of pH on either enzyme activity or reagent integrity, because the inactivation was assessed relative to enzyme that was also incubated at the appropriate pH, and DACM did not lose its ability to inactivate the enzyme at pH 7 following lengthy preincubations at pH 4–8 (see "Experimental Procedures"). If we assume that 1) the prototropic equilibrium is rapid

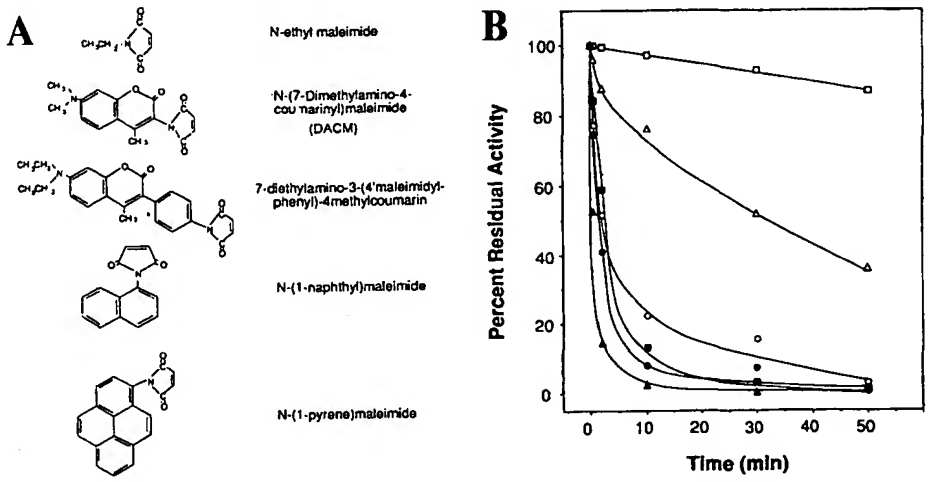

FIG. 2. **Inactivation of phosphoenolpyruvate carboxykinase by maleimide derivatives.** *Panel A*, chemical structures of the maleimide derivatives. *Panel B*, phosphoenolpyruvate carboxykinase (1 μM) was incubated at 0 °C in the presence of 10 mM TES (pH 7.0), 5% glycerol, 0.5 mM EDTA, 5% (v/v) acetone without maleimide (□) or plus a 2-fold molar excess of each maleimide: △, *N*-ethylmaleimide; ○, 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin; ●, DACM; ■, *N*-(1-naphthyl)maleimide; ▲, *N*-(1-pyrene) maleimide. Aliquots were taken at the indicated times to measure the rate of formation of oxaloacetate.

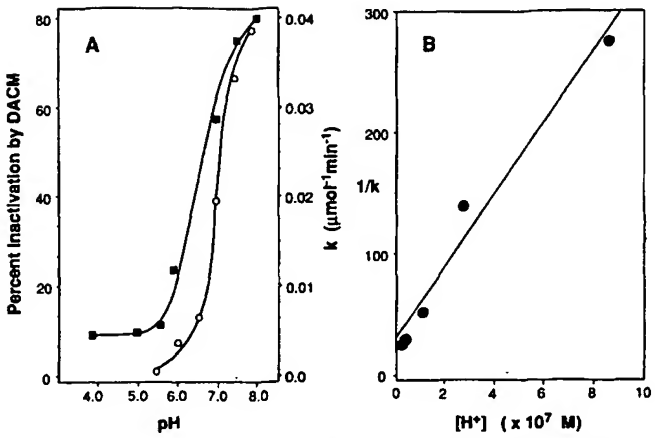

FIG. 3. **Dependence of rate of inactivation on pH.** *Panel A*, phosphoenolpyruvate carboxykinase was incubated with DACM or *N*-ethylmaleimide as described under "Experimental Procedures." *Left vertical axis*: the extent of inactivation of enzyme by equimolar DACM at each pH was determined after 5 min of incubation (■). *Right vertical axis*: the observed second order rate constants (determined as described under "Experimental Procedures") for the inactivation of phosphoenolpyruvate carboxykinase by *N*-ethylmaleimide are shown for each pH tested (○). *Panel B*, the reciprocals of the observed second order rate constants for the inactivation by *N*-ethylmaleimide (shown in *Panel A*) are plotted *versus* $[H^+]$. *K*, the ionization constant for the modified residue, was calculated by dividing the intercept by the slope (Equation 1, Ref. 26). The slope corresponds to a least squares fit of the data.

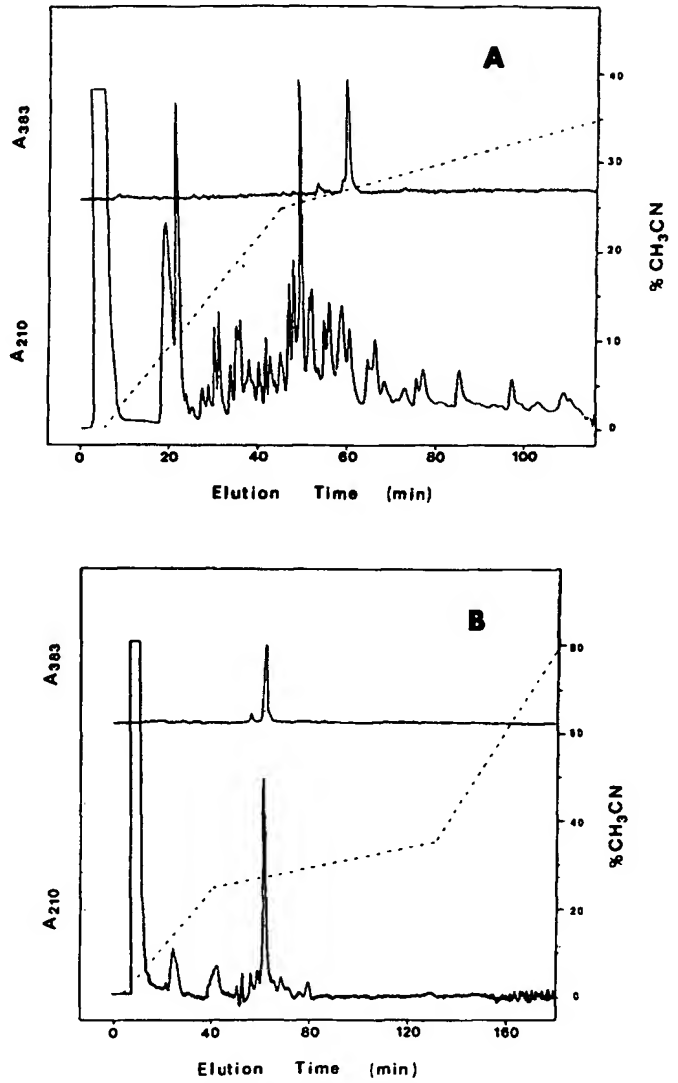

FIG. 4. **Purification of the modified peptide.** *Panel A*, digested modified enzyme was prepared and subjected to reversed-phase HPLC as described under "Experimental Procedures." The full-scale absorbance units are 0.25 at 210 nm and 0.03 (base-line offset 70%) at 383 nm. Fractions of 1 ml were collected, and the fractions corresponding to the middle of the single peak absorbing at 383 nm were pooled and lyophilized. *Panel B*, the partially purified peptide of *panel A* was dissolved in 10 mM ammonium acetate (pH 6.0), 4 M guanidine HCl and subjected to reversed-phase HPLC as described under "Experimental Procedures." The full-scale absorbance units are 0.2 at 210 nm and 0.1 (base-line offset 70%) at 383 nm. The yield of the purified peptide was not determined.

compared to the alkylation reaction, 2) the protonated Cys-SH is not alkylated (*cf.* Scheme 2*A* in Ref. 33), and 3) the same essential cysteine is modified by DACM and *N*-ethylmaleimide, then we can conclude that the apparent pK_a of this critical cysteine, shown below to be Cys-288, is approximately 7.0.

Purification of the Labeled Peptide—The reversed-phase HPLC peptide map of a tryptic digest of the modified enzyme is shown in Fig. 4*A*. The peptides were eluted with a triphasic linear gradient of acetonitrile as described under "Experimental Procedures"; however, no additional peptides were eluted in the third phase (35–80% acetonitrile). A single modified peptide was observed, as shown by the single symmetrical peak observed at 383 nm, eluting at approximately 59 min. Fractions corresponding to only the middle of the peak were pooled and repurified by reversed-phase HPLC (Fig. 4*B*). These procedures yielded a pure labeled peptide that was subjected to sequence analysis. The small peak visible at 383 nm that precedes the major peak may represent a breakdown product of DACM (it coelutes with free DACM), incomplete tryptic digestion, or minor labeling of a second cysteine residue.

Sequence of the Modified Peptide—The sequence of the purified peptide (Table II) was Tyr-Leu-Ala-Ala-Ala-Phe-Pro-Ser-Ala-*X*-Gly-Lys, where *X* was not identified. This sequence matches exactly that of Tyr-279 to Lys-290 of the amino acid sequence of phosphoenolpyruvate carboxykinase, which includes a cysteine residue at position 288 (46% of the distance from the amino terminus) (8). No phenylthiohydantoin derivative could be identified for the 10th cycle of sequencing, and no residues were detected after the 12th cycle of sequencing. These results unequivocally identify the essential hyperreactive cysteine of phosphoenolpyruvate carboxykinase as cysteine 288.

DISCUSSION

Despite the presence of 13 total cysteines, our results show that cysteine 288 can be labeled nearly completely by DACM without sacrificing specificity. Even at low temperature and a low molar excess of sulfhydryl reagent, the thiol was rapidly modified, resulting in the nearly complete loss of catalytic activity. The 1:1 correlation between inactivation and stoichiometry of labeling and the appearance of a single labeled peptide on reversed-phase HPLC peptide maps are evidence that only one cysteine was alkylated. Although we were unable to determine a second order rate constant for the reaction with DACM, it is clear that this thiol is unusually reactive.

The essential nature and pronounced reactivity of cysteine 288 raise questions concerning its microenvironment. The amino acid sequence surrounding the cysteine and the efficacy of the hydrophobic maleimide derivatives suggest that this thiol lies within a hydrophobic domain. Beale *et al.* (8) identified four hydrophobic regions in the phosphoenolpyruvate carboxykinase sequence: residues 138–223, 281–386, 398–460,

TABLE II

Sequence of the purified peptide

Approximately 250 pmol of the purified peptide was sequenced as described under "Experimental Procedures." The identity and yield (in pmol) of each phenylthiohydantoin derivative is listed for each cycle of sequencing. No residue was detected at position 10, and no additional residues were detected after the 12th cycle of sequencing. The identical sequence was obtained for a second sample of purified peptide that was generated from a different enzyme preparation.

Cycle number	Residue	Yield
		pmol
1	Tyr	174
2	Leu	183
3	Ala	196
4	Ala	194
5	Ala	192
6	Phe	33
7	Pro	118
8	Ser	98
9	Ala	113
10		
11	Gly	79
12	Lys	22

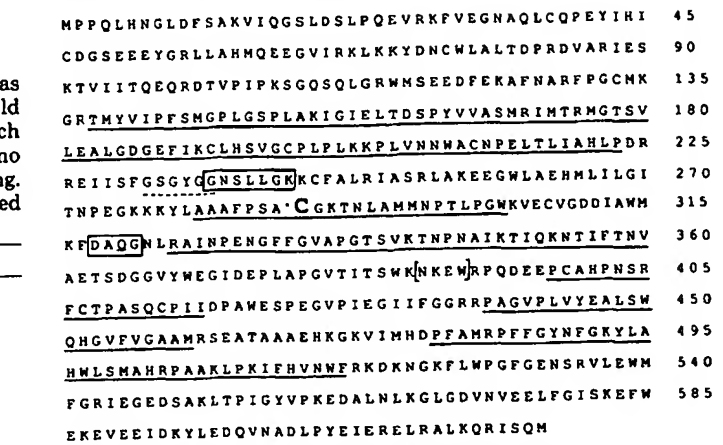

FIG. 5. **Amino acid sequence of phosphoenolpyruvate carboxykinase from rat liver cytosol (8).** Regions of interest are noted as follows: *, **bold**, cysteine 288; *dashed underline*, consensus sequence for a phosphoryl binding site common to many ATP-binding proteins (37); *boxed*, consensus sequences for phosphoryl binding sites of GTP-binding proteins (36); *bracketed*, consensus sequence for a guanine binding site, as discussed by Cook *et al.* (36); *solid underline*, extended hydrophobic regions (8).

and 479–517. Omitting some sequences of charged amino acids from the above stretches, the shorter hydrophobic regions illustrated in Fig. 5 are defined. Cysteine 288 lies within the hydrophobic sequence bounded by alanine 281 and tryptophan 303. DACM, in spite of its bulk, is clearly not prevented from rapidly modifying this thiol, nor are the larger or more nonpolar derivatives (Fig. 2), assuming of course that they are modifying the same cysteine. We conclude that the critical cysteine lies within a microenvironment of the enzyme that is relatively hydrophobic but that is nevertheless accessible to solvent. These results are consistent with previous documentation of hydrophobic domains of phosphoenolpyruvate carboxykinase. Hydrophobic interactions were thought to play a role in the binding of the enzyme to agarose-hexane-GTP, inasmuch as enzyme activity was also retained on columns of hexyl-, butyl-, and ethyl-agarose (9). Also, as was previously discussed, several fluorescent probes have been reported to bind within a hydrophobic environment of the mitochondrial carboxykinase and substrates protected against the resultant inactivation (30, 31). Thus, there is considerable evidence for a hydrophobic domain of the enzyme that is associated with catalytic activity.

Our data suggest that the hyperreactivity of cysteine 288 may be partially explained by a low pK_a. Because of the similarities of the pH dependence of enzyme inactivation by DACM and *N*-ethylmaleimide and because the inflection point of the pH dependence of inactivation by DACM is approximately 6.8 (Fig. 3*A*), we think it is reasonable, based on the previously discussed assumptions, to assign to cysteine 288 an apparent pK_a of 7.0, the pK_a determined with *N*-ethylmaleimide. We cannot, however, exclude the possibility that residues other than, or in addition to, Cys-288 are modified by *N*-ethylmaleimide and are contributing to the observed pH dependence of enzyme inactivation. The apparent pK_a for Cys-288 that we report is lower than the typical range of pK_a values of cysteinyl residues in proteins of 8–11 (34). One possible explanation for the low pK_a is that positively charged residues may lie in the vicinity of Cys-288 such that the thiol proton is more readily dissociated (due to stabilization of the resulting thiolate anion). Because almost all of the enzyme's substrates are negatively charged at neutral pH, one might expect to find positively charged amino acids within the active site.

Based upon its characteristics and location, cysteine 288 is very likely the essential cysteine previously characterized by Carlson *et al.* (17) using Nbs_2. That cysteine was estimated to be approximately 44% of the distance from the amino terminus; Cys-288, at 46% of the distance, is the closest of all of the 13 cysteines to this location. The essential nature of Cys-288 was apparent from our studies; the residual activity of the modified enzyme was always low, and the enzyme's abilities to catalyze the formation of oxaloacetate, phosphoenolpyruvate, or pyruvate were destroyed in parallel. Furthermore, we always observed 1 mol of thiol modified per mol of inactivated enzyme. Inactivation of phosphoenolpyruvate carboxykinase by equimolar Nbs_2 has similar effects; the decreases in the enzyme's three activities occur in parallel, the residual activity is very low, and the relationship between fractional inactivation and fractional modification is 1:1 (17). The rates of enzyme inactivation by equimolar Nbs_2 and DACM were similar in that in both cases inactivation was almost too rapid to measure. Given that a single hyperreactive essential cysteine is rapidly modified by both Nbs_2 and DACM, it is certainly not unreasonable to expect that each reagent may be modifying the same residue. Comparison of the abilities of substrates to protect against enzyme inactivation by DACM and Nbs_2 provides further evidence that the same cysteine is modified by both reagents. Metal nucleotides afforded the best protection against inactivation and modification, whereas phosphoenolpyruvate protected against the rate, but not the extent, of modification. Oxaloacetate was the least effective substrate tested in both studies. The only significant difference in the substrate protection studies was that 1–2 mol of thiol/mol of enzyme were still modified by equimolar Nbs_2 after extended incubation in the presence of IDP or ITP, but because the residual activity of the modified enzyme was still quite high, the thiols modified by Nbs_2 under those conditions had to be other than the essential thiol. In contrast, we could detect no incorporation of DACM in the presence of GDP even after extended incubation. It is possible that the size or hydrophobic nature of DACM increases its specificity such that thiols other than cysteine 288 are not readily modified. The sum of these data provides strong evidence that the essential hyperreactive cysteine that is modified by equimolar Nbs_2 is, in fact, cysteine 288.

If this supposition is correct, our knowledge of cysteine 288 can be summarized as follows. 1) Cysteine 288 is an essential

residue of cytosolic phosphoenolpyruvate carboxykinase. Although we would be hesitant to draw such a conclusion based solely on inactivation caused by bulky modifying groups, the enzyme is equally inactive when the nitrobenzoate attached to the essential cysteine is replaced by a CN moiety (17). Moreover, enzyme that has been inactivated by Nbs_2 can be almost completely reactivated by excess dithiothreitol (17). 2) Cysteine 288 is unusually reactive. The ability of the thiol to be rapidly and exclusively labeled in the presence of 12 other cysteines attests to its hyperreactivity. 3) The apparent pK_a of the critical cysteine is approximately 7.0. 4) The thiol resides within a hydrophobic microenvironment. 5) Although activation by free Mn^{2+} is dependent upon the proper oxidation state of the enzyme's cysteine residues (16, 21, 35), Cys-288 is probably not involved in binding the activating free divalent cation, because Mn^{2+} or Mg^{2+} does not protect against inactivation by DACM. 6) Cysteine 288 is proximal to a second cysteine, with which it can be readily induced to form a cystine disulfide (17).

Does this critical thiol lie within the active site? To date, no substrate binding sites of phosphoenolpyruvate carboxykinase have been identified, so we are unable to determine whether cysteine 288 is, indeed, an active site residue; however, our data are certainly consistent with such a conclusion. All of the substrates tested protect against modification of the cysteine, and nucleotides offer almost complete protection. Furthermore, cysteine 288 is located near several consensus sequences for nucleotide binding sites (Fig. 5). Although the enzyme has a specific requirement for guanosine or inosine nucleotides, it does not share a great deal of homology with other GTP-binding proteins. It has been suggested that residues 388–391 comprise a consensus sequence for a guanine-binding site (36). In addition, the glycine-rich region that extends from residues 232 to 243 may comprise a consensus sequence for a phosphoryl binding site common to both ATP- and GTP-binding proteins (36, 37). Cook *et al.* (36) also pointed out a second consensus sequence for a phosphoryl binding site common to GTP-binding proteins, extending from Asp-318 to Gly-321. Cys-288 lies between these two putative phosphoryl binding domains, approximately 30–40 residues from each. Although the role of this critical cysteine is not known, it is improbable that its sole function is to bind free divalent cations, as discussed above, or to participate in phosphoryl transfer, because thiol modification also prevents the decarboxylation of oxaloacetate to pyruvate, a reaction that does not include phosphoryl transfer. Regardless of its role cysteine 288 is unusually reactive, and its integrity is essential for catalysis.

Acknowledgment—We wish to thank Dr. Bob Cassell of the Veterans Administration Medical Center of Memphis for his excellent technical assistance in peptide sequencing.

REFERENCES

1. Weldon, S. L., Hod, Y., Savon, S. P., Shedroff, A. H., Cook, J. S., and Hanson, R. W. (1987) *Fed. Proc.* **46**, 2181
2. Ballard, F. J., and Hanson, R. W. (1969) *J. Biol. Chem.* **244**, 5625–5630
3. Hod, Y., Utter, M. F., and Hanson, R. W. (1982) *J. Biol. Chem.* **257**, 13787–13794
4. Ballard, F. J. (1971) *Biochim. Biophys. Acta* **242**, 470–472
5. Hammond, K. D., and Balinsky, D. (1978) *Int. J. Biochem.* **9**, 199–211
6. Holten, D. D., and Nordlie, R. C. (1965) *Biochemistry* **4**, 723–731
7. Gallwitz, W. E., Jacoby, G. H., Ray, P. D., and Lambeth, D. O. (1988) *Biochim. Biophys. Acta* **964**, 36–45
8. Beale, E. G., Chrapkiewicz, N. B., Scoble, H. A., Metz, R. J., Quick, D. P., Noble, R. L., Donelson, J. E., Biemann, K., and Granner, D. K. (1985) *J. Biol. Chem.* **260**, 10748–10760
9. Colombo, G., Carlson, G. M., and Lardy, H. A. (1978) *Biochemistry* **17**, 5321–5329
10. Utter, M. F., Kurahashi, K., and Rose, I. A. (1954) *J. Biol. Chem.* **207**, 803–819
11. Hebda, C. A., and Nowak, T. (1982) *J. Biol. Chem.* **257**, 5503–5514
12. Bentle, L. A., and Lardy, H. A. (1976) *J. Biol. Chem.* **251**, 2916–2921
13. Cheng, S. C., and Cheng, R. H. C. (1972) *Arch. Biochem. Biophys.* **151**, 501–511
14. Reeves, J. B. (1975) *Diss. Abstr. Int. B Sci. Eng.* **36**, 2199-B
15. Ballard, F. J., and Hopgood, M. F. (1976) *Biochem. J.* **154**, 717–724
16. Brinkworth, R. I., Hanson, R. W., Fullin, F. A., and Schramm, V. L. (1981) *J. Biol. Chem.* **256**, 10795–10802
17. Carlson, G. M., Colombo, G., and Lardy, H. A. (1978) *Biochemistry* **17**, 5329–5338
18. Barns, R. J., and Keech, D. B. (1972) *Biochim. Biophys. Acta* **276**, 284–296
19. Barns, R. J., and Keech, D. B. (1968) *Biochim. Biophys. Acta* **159**, 514–526
20. Barns, R. J., and Keech, D. B. (1972) *Biochim. Biophys. Acta* **289**, 410–419
21. Makinen, A. L. (1984) Ph.D. dissertation, University of Notre Dame
22. Goto, Y., Shimizu, J., and Shukuya, R. (1980) *J. Biochem. (Tokyo)* **88**, 1239–1249
23. Yamamoto, K., Sekine, T., and Kanaoka, Y. (1977) *Anal. Biochem.* **79**, 83–94
24. Haugland, R. P. (1985) *Molecular Probes: Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Eugene, OR
25. Colombo, G., and Lardy, H. A. (1981) *Biochemistry* **20**, 2758–2767
26. Lindley, H. (1962) *Biochem. J.* **82**, 418–425
27. Allen, G. (1981) *Laboratory Techniques in Biochemistry and Molecular Biology: Sequencing of Proteins and Peptides*, pp. 28–31, Elsevier/North-Holland Biomedical Press, Amsterdam
28. Hewick, R. M., Hunkapillar, M. W., Hood, L. E., and Dreyer, W. J. (1981) *J. Biol. Chem.* **256**, 7990–7997
29. Seyer, J. M., Mainardi, C., and Kang, A. H. (1980) *Biochemistry* **19**, 1583–1589
30. Silverstein, R., Rawitch, A. B., and Grainger, D. A. (1979) *Biochem. Biophys. Res. Commun.* **87**, 911–918
31. Silverstein, R., Lin, C-C., and Rawitch, A. B. (1980) *J. Biol. Chem.* **255**, 1374–1379
32. Nelson, D. L., and Silverstein, R. (1982) *Biochem. Biophys. Res. Commun.* **108**, 696–699
33. Knowles, J. R. (1976) *CRC Crit. Rev. Biochem.* **4**, 165–173
34. Tanford, C. (1962) *Adv. Protein Chem.* **17**, 69–165
35. Johnston, J. B. (1970) Ph.D. dissertation, University of Wisconsin
36. Cook, J. S., Weldon, S. L., Garcia-Ruiz, J. P., Hod, Y., and Hanson, R. W. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 7583–7587
37. Hannink, M., and Donoghue, D. J. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 7894–7898